

Tecnologias de visualização sonora num contexto artístico de performance musical

Pedro Emanuel de Oliveira Santos

Licenciado em Produção e Tecnologias da Música
pelo Instituto Politécnico do Porto

Dissertação submetida para satisfação parcial dos requisitos
do grau de Mestre em Multimédia

Dissertação realizada sob a orientação do Professor Auxiliar

Aníbal João de Sousa Ferreira

da Faculdade de Engenharia da Universidade do Porto

Porto, Outubro 2009

Agradecimentos

Numa altura como a presente, em que culminamos este processo académico, torna-se essencial fazer uma menção especial e agradecer a todos quantos de uma forma ou de outra contribuíram para o desenvolvimento desta dissertação, tornando assim possível a sua elaboração.

A um nível académico, gostaria em primeiro lugar de endereçar uma palavra de apreço e gratidão ao Professor Auxiliar Aníbal Ferreira, enquanto orientador desta dissertação, pela sua orientação e apoio ao longo de todo o tempo de desenvolvimento da mesma. Sem a sua exigência científica, os seus conselhos e a sua leitura crítica do trabalho desenvolvido, o resultado final teria sido certamente bem diferente. Gostaria também de agradecer ao Prof. Luís Gustavo Martins pela leitura crítica do trabalho e de algumas sugestões efectuadas ainda numa fase inicial. Gostaria ainda de agradecer ao Prof. Eurico Carrapatoso, enquanto Director do Mestrado em Tecnologia Multimédia e, mais recentemente, do Mestrado em Multimédia, pela sua pronta colaboração, sempre que solicitada, em particular no processo de reingresso e adequação.

Num contexto profissional, gostaria de agradecer aos meus colegas docentes na ESMAE/IPP, em particular do Curso de Música - variante em Produção e Tecnologias da Música, pela sua contribuição directa, através de conselhos ou orientações dadas em situações pontuais, mas sobretudo pela partilha de experiências a que diariamente sou sujeito.

E, por último, mas certamente não por isso menos sentido, num contexto muito pessoal, gostaria de agradecer imenso à Ruth, minha esposa, pelo amor e apoio com que sempre me premiou e até por algum sacrifício pessoal a que se sujeitou de modo a me permitir despendir um tempo e energia consideráveis neste trabalho; aos meus filhos, Rafael e Beatriz "apenas" por serem como são, pelo amor e alegria contagiante que transmitem e que ilumina as nossas vidas; a todos os restantes membros da minha família, pelo sólido suporte afectivo que sempre me proporcionaram; e a Deus, sem o qual nada seria possível.

Prefácio

A presente dissertação foi desenvolvida tendo por base um interesse prévio pela área abordada. O âmbito de estudo foi formulado ainda no contexto do Mestrado em Tecnologia Multimédia, tendo sido aí desenvolvido um trabalho inicial de investigação e exploração. No ano lectivo de 2008/2009 foi feito o reingresso e adequação ao novo Mestrado em Multimédia com o intuito de entregar e defender a presente dissertação. Dada a diferença temporal entre estes dois momentos, o da formulação inicial do assunto a abordar, e o da entrega final do trabalho, alguns dos aspectos mais "inovadores" que definimos inicialmente têm-se vindo a generalizar, embora que de forma ténue, pelo que consideramos que a validade da problemática contemplada na presente dissertação continua ainda hoje extremamente válida.

Resumo

A presente dissertação de mestrado centra-se na exploração de ferramentas tecnológicas que potenciem a visualização sonora num contexto criativo/artístico de *performance* musical. Através da atribuição de correlações entre parâmetros do domínio sonoro e visual, pretende-se obter da acção musical uma correspondente reacção visual.

As implementações abordadas nesta tese fazem uso de tecnologias como a análise de áudio digital, o MIDI, a síntese sonora e o Open Sound Control como ponto de partida para a visualização musical. Para a construção visual é utilizado o OpenGL (2D/3D), incluindo tecnologias de geração e processamento de texturas (*shaders* programáveis) de modo a tirar partido do desempenho das placas gráficas actuais. Para além das correlações audiovisuais sugeridas, de forma a potenciar a expressividade do sistema visual, é também contemplada a possibilidade do controlo directo deste por um “instrumentista visual” através de dispositivos como *gamepads*, mesas digitalizadoras, ou equipamentos MIDI. Tendo em conta a eventual expansão do sistema proposto é ainda sugerida a possibilidade de recurso a sistemas de processamento distribuído através de redes informáticas assentes na arquitectura protocolar TCP/IP, abordando assuntos como a transmissão de áudio, MIDI, OSC, geometria e texturas entre computadores.

Os sistemas propostos nesta dissertação foram concretizados através de ambientes de programação visual de alto nível – o Reaktor para a síntese sonora e o Max/MSP/Jitter para a restante programação de controlo e geração de imagem – mais próximos dos artistas pelo seu maior grau de acessibilidade, e mais apropriados a uma programação com necessidades de *feedback* em tempo-real.

Abstract

This current Master thesis revolves around the exploration of technological tools capable of empowering the field of Sound Visualization, in a creative/artistic context of musical performance. Through the assignment of correlations between parameters in the sound and visual domains, it is intended to obtain a visual reaction to the corresponding musical actions.

The implementations explored in this thesis make use of technologies such as digital audio analysis, MIDI, and sound synthesis in conjunction with the Open Sound Control protocol as a starting point towards musical visualization. To build the visual representation, we use the OpenGL API (2D/3D), including texture generation and processing technologies (programmable shaders) in order to take advantage of current graphics cards' performance. In addition to the audiovisual correlations suggested, in order to maximize the visual system's expressivity, it is also considered the possibility of its direct control by a "visual artist/performer" through devices such as gamepads, pen tablets or MIDI equipments. Considering the eventuality of the proposed system's expansion, we also suggest the ability to rely on distributed processing systems through computer networks based on the TCP/IP architecture, exploring fields such as digital audio, MIDI, OSC, geometry and texture communication between computers.

The proposed systems in this thesis were developed on high-level graphical programming environments – Reaktor for the sound synthesis, and Max/MSP/Jitter for the remaining control and image processing programming – closer to the artists due to their higher accessibility, and more appropriate to a programming workflow with real-time feedback requirements.

Índice

Agradecimentos	2
Prefácio	3
Resumo	4
Abstract.....	5
Índice.....	6
Introdução.....	9
Capítulo I - O Som, a Imagem, a Arte e a Técnica.....	13
1.1 O Som	14
1.1.1 A audição.....	15
1.1.2 O Fenómeno Físico	16
1.2 A Imagem	17
1.2.1 A visão	18
1.2.2 O Fenómeno Físico	19
1.3 Estímulo e Sensação	20
1.4 Correspondências audiovisuais?.....	21
1.4.1 A cor e as notas musicais	22
1.4.2 A cor e o timbre	23
1.4.3 A cor e o modo.....	23
1.4.4 Associações a outras características da cor	24
1.4.5 O movimento e a representação temporal.....	24
1.4.6 O som e a forma.....	25
1.4.7 Um nível superior	25
1.4.8 Opiniões contrárias	26
1.5 Instrumentos audiovisuais	28
1.5.1 Século XVIII.....	28
1.5.2 Século XIX	29
1.5.3 Século XX.....	29
1.5.4 Presente	31
1.6 A Música e as Artes Visuais.....	33
1.6.1 Pintura.....	33
1.6.2 Cinema e Vídeo.....	35
Capítulo II - Oportunidades tecnológicas	39
2.1 O Áudio Digital	40
2.1.1 Amostragem e Quantização.....	40

2.1.2	<i>Análise, classificação e descrição de conteúdos áudio</i>	41
2.2	O MIDI	45
2.2.1	<i>Estrutura de uma mensagem MIDI</i>	46
2.2.2	<i>Canais e Mensagens MIDI</i>	46
2.3	O OpenSound Control	49
2.3.1	<i>Resumo das características do OSC</i>	49
2.3.2	<i>Arquitectura de Rede do OSC</i>	50
2.3.3	<i>Implementação do protocolo OSC</i>	50
2.4	A Síntese Sonora	52
2.4.1	<i>Breve perspectiva histórica</i>	52
2.4.2	<i>A síntese sonora e o computador</i>	54
2.4.3	<i>Componentes de um sintetizador</i>	55
2.4.4	<i>Métodos de Síntese Sonora</i>	58
2.5	A Imagem Digital	60
2.5.1	<i>Gráficos Bitmap</i>	60
2.5.2	<i>OpenGL</i>	61
2.5.3	<i>Vídeo e Animação</i>	64
2.6	Exemplos de Visualização Sonora	66
2.7	Aplicações e ambientes de programação	70
2.7.1	<i>Análise dos ambientes de programação</i>	70
2.7.2	<i>A escolha dos ambientes de programação</i>	73
Capítulo III - Implementação		75
3.1	Introdução	76
3.2	Comunicação entre aplicações e Processamento distribuído	80
3.2.1	<i>Transmissão de MIDI e Áudio entre aplicações locais e em rede</i>	80
3.2.2	<i>Processamento distribuído</i>	83
3.3	Visualização de Áudio Digital	87
3.3.1	<i>Música pré-gravada</i>	87
3.3.2	<i>Performance ao vivo</i>	88
3.3.3	<i>Recursos de análise disponíveis</i>	89
3.3.4	<i>Programação</i>	92
3.4	Visualização de MIDI	94
3.4.1	<i>Controladores e Conversores MIDI</i>	94
3.4.2	<i>Mensagens MIDI e possibilidades de controlo</i>	95
3.4.3	<i>Esquemas de ligações</i>	96
3.4.4	<i>Programação</i>	98
3.5	Visualização de Síntese Sonora	100
3.5.1	<i>Componentes do Sistema Sonoro</i>	100

3.5.2	<i>Comunicação entre os sistemas através de OSC.....</i>	<i>104</i>
3.5.3	<i>Relações estabelecidas entre o Sintetizador Sonoro e o Visual</i>	<i>107</i>
3.6	O Sistema Visual	109
3.6.1	<i>Conceito.....</i>	<i>109</i>
3.6.2	<i>Áudio Digital</i>	<i>113</i>
3.6.3	<i>MIDI</i>	<i>115</i>
3.6.4	<i>Síntese Sonora</i>	<i>117</i>
3.7	Métodos de Controlo	128
3.7.1	<i>Controlo do Sintetizador Sonoro.....</i>	<i>128</i>
3.7.2	<i>Controlo do Sistema Visual</i>	<i>132</i>
	Capítulo IV - Conclusão	138
4.1	<i>Avaliação e análise crítica</i>	<i>139</i>
4.2	<i>Futuros desenvolvimentos</i>	<i>142</i>
	Bibliografia	144

Introdução

Ao longo do tempo, muitos artistas visuais têm acreditado que a música oferece um modelo ao qual a arte visual deveria aspirar: uma forma de arte pura e abstracta que vai para além da realidade perceptível. Uma das designações utilizadas para definir o relacionamento entre as artes visuais e a música foi sugerida em 1912 pelo crítico de arte Roger Fry e tem sido utilizada até aos dias de hoje: “Visual Music” [1]. No entanto, este termo não é único. A utilização sincronizada entre música e imagens abstractas é também muitas vezes designada por “Ocular Music”, “Color Music”, “Mobile Color” ou “Lumia” [2]. Os artistas têm utilizado estas expressões para definir os seus esforços em integrar os sentidos através da arte. Apesar de pouco difundida, esta ideia de relacionamento dos sentidos, proporcionando a fusão da música com as artes visuais é já remota, tendo sido abordada por famosos pensadores, matemáticos e cientistas como Aristóteles, Pitágoras, Leonardo da Vinci ou Isaac Newton. Estes e muitos outros reflectiram na possibilidade de existência de correspondências entre o fenómeno sonoro e o fenómeno luminoso.

A propósito da falta de conhecimento generalizada relativamente à chamada *Visual Music*, Adrien Bernard Klein afirma em 1927 no seu livro “*Color-Music: the Art of Light*” [3], que “*é um facto estranho que quase todos quantos constroem um órgão de luz estejam erradamente convencidos de que eles, ou elas, são o primeiro mortal a tentar fazê-lo.*” Estranhamente, até aos nossos dias esta percepção não tem sido consideravelmente alterada; ainda hoje esta área continua pouco divulgada e o extenso trabalho de personalidades como Louis-Bertrand Castel, Alexander Wallace Rimington, Thomas Wilfred, Oskar Fischinger ou James Whitney continua amplamente desconhecido. Uma das pessoas que mais se destacou na tentativa de inverter esta tendência foi William Moritz, ao ocupar uma grande parte da sua vida a pesquisar sobre o assunto, tendo escrito imensas publicações sobre pessoas tão importantes como Oskar Fischinger, James Whitney, Jordan Belson, Hy Hirsh, Mary Ellen Bute, entre outros. Nos últimos anos da sua vida, Moritz trabalhou em conjunto com Elfriede Fischinger, viúva de Oskar Fischinger, para dar a conhecer a obra deste a um público mais alargado, esforço que se traduziu em 2004 na publicação da biografia definitiva de Fischinger, “*Optical Poetry*” [4]. Pela sua enorme contribuição para a documentação e divulgação da *Visual Music*, consideramos oportuna esta referência, em forma de homenagem, a William Moritz.

O fenómeno de sinestesia tornou-se numa importante influência para o desenvolvimento da *Visual Music*. Este pode ser definido por uma unidade dos sentidos e, por consequência, uma síntese das artes [5]. A teoria da sinestesia defende que determinada percepção sensorial pode induzir percepções em outros sentidos. Assim, torna-se possível a alguém com esta percepção sinestésica, por exemplo, perceber uma certa cor sempre que ouve determinada nota. É de referir que a música, dada a sua influência nas outras formas de arte, tem sido considerada por muitos como a fonte por excelência deste fenómeno. Apesar dele não estar restrito ao relacionamento dos sentidos auditivo e visual, esta tem sido ao longo do tempo a associação mais recorrente. Sendo o som e a imagem fenómenos físicos totalmente distintos e manifestando-se por isso mesmo de forma distinta, torna-se inútil tentar comprovar objectivamente que existem relações entre determinadas características sonoras e visuais. Podemos no entanto afirmar com algum grau de certeza que haverá alguns relacionamentos mais prováveis do que outros. Em muitas situações da nossa vida quotidiana,

servimo-nos de termos provenientes de outros sentidos, como *quente* ou *suave*, para caracterizar uma cor, *alto* ou *sombrio* para caracterizar determinado som. Este facto sugere que, mesmo de forma inconsciente, possuímos uma tendência para relacionar as diferentes percepções dos nossos sentidos, construindo correspondências entre eles permanentemente.

Acreditamos que a combinação dos universos sonoro e o visual potencia uma forma única de comunicação mas, sobretudo, de expressão; a sua fusão dá origem a algo mais que o mero somatório das partes. Michel Chion, no seu livro *Audio-Vision* [6] refere que este fenómeno de “valor acrescentado” tem uma incidência particular no caso da combinação entre som e imagem, pelo estabelecimento de uma imediata e necessária relação entre algo que ouvimos e algo que vemos quando estes eventos acontecem em simultâneo. Estas duas formas de percepção acabam por se complementar, influenciando-se reciprocamente, cada uma emprestando à outra as suas respectivas características por contaminação e projecção. O som faz-nos ver a imagem de uma forma diferente do que se vista sozinha, do mesmo modo que a imagem permite-nos ouvir o som de uma forma bem distinta do que se ouvido “no escuro”.

Nos dias que correm, a imagem possui um papel cada vez mais importante na indústria musical. Desde que o vídeo musical *Video Killed the Radio Star* dos *The Buggles* foi utilizado na estreia da MTV (Music Television), a 1 de Agosto de 1981, que qualquer *single* de sucesso é obrigatoriamente acompanhado por um vídeo musical. Paralelamente, em actuações ao vivo, montam-se grandes telas de vídeo para acompanhar os concertos, sendo a apresentação baseada em câmaras de vídeo a captar a actuação dos músicos, combinadas muitas vezes com outro tipo de imagens, abstractas ou não. Também no campo da audição musical através das tradicionais aplicações de leitura de música, como o *iTunes*, o *Winamp* ou o *Windows Media Player*, todas elas possuem visualizadores de música incluídos, bem como a possibilidade de extensão desta funcionalidade através do desenvolvimento de *plug-ins* por terceiros. Nesta área, têm aparecido ideias interessantes na exploração do aspecto visual, mas no que concerne à relação criada entre a música ouvida e a visualização gerada, não existe, infelizmente, muita variedade. Parâmetros como a cor, a posição, o tamanho, o movimento, entre outros, não costumam possuir qualquer correlação com o som e a música, sendo seleccionados valores de acordo com algoritmos pré-concebidos ou de forma aleatória/semi-aleatória. Técnicas de detecção de características sonoras têm sido desenvolvidas nos meios académicos, normalmente com finalidades de análise objectiva dos mesmos, mas, infelizmente, estas raramente têm sido aplicadas a este tipo de visualizadores com um carácter criativo/artístico. A sua utilização potencialmente poderia tornar os visualizadores mais interessantes, pois a correlação entre som e imagem resultante seria muito maior.

O artista visual Paul Friedlander, conhecido pelas suas esculturas de luz, divide a *Visual Music* em três categorias distintas [7]:

1. uma forma de conversão rígida e directa da música em imagens, utilizando um conjunto de regras para o relacionamento audiovisual e interpretadas através de *hardware* ou *software* adequado;

2. uma forma de expressão da música através de imagens, na qual se configura como fundamental o envolvimento activo de um artista que interprete a música e encontre a melhor forma de a expressar visualmente;
3. uma forma de expressão visual que não possua uma relação directa com a música, embora possa ser comparada a ela em termos funcionais e influenciada por ela a um nível estrutural.

A abordagem por nós seguida nesta tese é coincidente com a primeira interpretação, na medida em que, através da utilização de vários meios tecnológicos, procuramos obter um nível de discretização dos parâmetros musicais suficiente para o desejado relacionamento com o domínio visual. Por outro lado, revemo-nos igualmente na segunda definição, pois temos a consciência de que a interpretação de qualquer forma de expressão artística por um ser humano possibilitará obter outro tipo de resultados, a um nível semântico já mais elevado. Somos da opinião que estas duas interpretações da *Visual Music* não têm de ser julgadas como radicalmente opostas e inconciliáveis, mas, pelo contrário, devem ser encaradas como complementares, tal como pretendemos demonstrar no nosso trabalho.

Nesta dissertação, pretendemos investigar/explorar os recursos tecnológicos que potenciem o desenvolvimento de um sistema baseado em *software* e *hardware* que permita a exploração de relacionamentos entre som e imagem, num contexto prático de *performance* musical em tempo-real. Sendo os principais *interfaces* de controlo do sistema proposto essencialmente musicais (áudio digital, controlador MIDI, sintetizador sonoro) e sendo a visualização uma consequência directa das opções musicais, podemos considerar que esta tese se enquadra no âmbito da visualização sonora, não com fins de medição objectiva de parâmetros, mas com uma motivação essencialmente estética/artística. Os relacionamentos entre parâmetros sonoros e parâmetros visuais por nós sugeridos no decorrer destas páginas não passam de isso mesmo, sugestões; apesar de termos tido algum critério na atribuição dessas relações, não possuímos o intuito explícito de as suportar teoricamente. Estando esta dissertação inserida num contexto essencialmente tecnológico, não pretendemos de forma alguma investigar e tirar conclusões objectivas sobre as possibilidades de correlação entre o que ouvimos e o que vemos. Essa motivação serviu e continuará a servir de base para outras publicações, em que, através da investigação de estudos relacionados com a percepção sonora e visual e através da construção de modelos básicos de correlação entre som e imagem, realização de inquéritos e respectivo tratamento estatístico dos dados obtidos, fosse possível chegar a alguma conclusão [8],[9]. A abordagem por nós preconizada aproxima-se mais da postura de Fischinger [4], que considerava que o mapeamento de correspondências entre parâmetros sonoros e visuais possui um carácter profundamente subjectivo e pessoal. Assim, o próprio mapeamento de relações entre estes dois meios constitui já à partida uma oportunidade para a expressão artística, opinião que funciona como premissa fundamental para esta tese.

Capítulo I - O Som, a Imagem, a Arte e a Técnica

1.1 O Som

Do ponto de vista da linguagem, todo o mundo audível que nos rodeia – a *audiosfera*, segundo Cloutier [10] – é constituída por 3 componentes essenciais:

1. *Fonte sonora*: emissor que gera informação acústica;
2. *Som*: meio de comunicação entre emissor e receptor;
3. *Audição*: captação e processamento do som por parte do receptor.

É a audição quem fornece um significado ao som; sem esta o som não passaria de um fenómeno físico de vibração de partículas num determinado meio. A audição é efectuada pelo ouvido mas é o cérebro o responsável pela descodificação sonora, ou seja, pela atribuição de um determinado significado ao fenómeno físico.

A linguagem áudio é temporal e linear. Cada som tem associada uma duração própria. Podemos quantificar a duração de um som através de uma unidade de tempo ou através de figuras rítmicas de notação, no caso do som estar inserido num contexto musical. Ao contrário da linguagem visual, onde a mensagem pode estar distribuída em termos espaciais (ex: folha de papel), na linguagem sonora a mensagem está sempre organizada de forma sequencial ao longo do tempo. Os signos sonoros são transmitidos um após outro, como as palavras num texto. Devido a esta dependência exclusivamente temporal, podemos considerar o fenómeno sonoro como unidimensional. O ritmo, ou seja, a organização de eventos sonoros ao longo do tempo, torna-se assim um dos elementos fundamentais desta linguagem pois estrutura o discurso e delimita as durações dos eventos sonoros.

O ouvido, responsável pela captação desta linguagem, está envolvido por uma atmosfera sonora. Eis uma diferença fundamental entre a audição e a visão: o ouvido capta sons provenientes de todas as direcções, proporcionando uma verdadeira imersão. No entanto, apesar desta característica, o processamento do som por parte do cérebro permite a selectividade, ou seja, a escolha de determinada informação da atmosfera sonora, filtrando a restante em termos de descodificação e atribuição de um significado.

Das várias formas de comunicação efectuadas neste meio, a palavra falada é a mais óbvia uma vez que possui à partida um determinado significado associado. No entanto, a música - que pode ser considerada como uma combinação organizada de sons - pode tornar-se também uma forma de comunicação, apesar do seu carácter ser inerentemente abstracto. A estruturação musical é efectuada segundo uma linguagem própria a ela associada e possui implicações artísticas para além do mero aspecto comunicacional. Para o propósito específico desta tese, esta última, a música, será a categoria mais visada.

1.1.1 A audição

O ouvido é constituído por 3 partes básicas - o ouvido externo, o ouvido médio, e o ouvido interno. Cada parte possui uma função específica na interpretação sonora. O ouvido externo tem a função de captar o som e de o encaminhar por um canal até ao ouvido médio. O ouvido médio, tem a função de transformar a pressão sonora em vibrações internas através da sua estrutura óssea, com o objectivo de amplificar o som e enviá-las para o ouvido interno. A função do ouvido interno é a de transformar a energia das ondas de pressão em impulsos nervosos que são posteriormente transmitidos ao cérebro, onde são processados [11].

Ouvido externo

O ouvido externo consiste numa formação cartilaginosa cheia de contornos em forma de funil, a que damos o nome de pavilhão da orelha, e no canal auditivo, com forma cilíndrica, cujo comprimento é de aproximadamente 25 mm. O ouvido externo capta o som, processo no qual o canal auditivo, amplifica as ondas sonoras duas a três vezes. O canal auditivo termina no tímpano, membrana resistente e bem esticada que capta a variação de pressão proveniente do canal auditivo, vibrando em conformidade.

Ouvido médio

O ouvido médio é uma cavidade cheia de ar, consistindo fundamentalmente em 3 pequenos ossos interligados: o martelo, a bigorna e o estribo. Estes agem como amplificadores das vibrações da onda sonora. A força de vibração da bigorna é aproximadamente 15 vezes maior do que a do tímpano. Esta característica aumenta a nossa possibilidade de ouvir o mais fraco dos sons. Os músculos do ouvido médio modificam o desempenho deste sistema de alavancas, actuando como uma unidade amplificadora, além de actuarem como um controlador de volume, pois protegem o ouvido contra sons excessivamente altos diminuindo as vibrações que chegam do tímpano.

Ouvido interno

A janela oval efectua a comunicação entre o ouvido médio e o ouvido interno. O tamanho da Janela Oval é de 15 a 30 vezes menor do que o tímpano. Esta concentração de forças permite a amplificação das vibrações sonoras na mesma proporção. O resultado conjunto de todos os mecanismos de amplificação permite que a onda sonora possa ser amplificada até mais de 800 vezes antes que o líquido do ouvido interno entre em movimento.

O ouvido interno é constituído pela cóclea, pelos canais semicirculares, e pelo nervo auditivo. Os canais semicirculares não têm qualquer função na audição; servem simplesmente para medir a sensação de equilíbrio do nosso corpo. A cóclea é um órgão em forma de caracol que, se esticada teria aproximadamente 3 cm. Esta encontra-se cheia de um líquido e a sua superfície interna está alinhada com cerca de 30.000 células nervosas com o nome de cílios que executam a função mais crítica da nossa capacidade de ouvir. Elas

possuem uma sensibilidade natural a uma determinada frequência de vibração. Quando a frequência da onda de pressão corresponde com a dos cílios, estes irão ressoar com uma grande amplitude de vibração, enviando assim impulsos eléctricos que passam ao longo do nervo auditivo para o cérebro. Através de um processo que ainda não é compreendido na totalidade, o cérebro é capaz de interpretar as características sonoras através da reacção a estes mesmos impulsos nervosos.

1.1.2 O Fenómeno Físico

Em termos físicos, o som é um fenómeno que possui uma natureza vibratória. É propagado num determinado meio através de ondas de pressão. A velocidade de propagação sonora no ar, meio mais comum, é de aproximadamente 340 m/s. É importante referir que a energia sonora é transmitida de um ponto a outro sem que as partículas do meio se desloquem em conjunto com a onda [12]. As principais características do som são a *intensidade*, a *altura* e o *timbre*.

Intensidade: a intensidade sonora está relacionada com a amplitude da onda sonora, ou seja, com a quantidade de energia por ela transportada. Quanto maior for esta amplitude, maior será a sua intensidade. A unidade de medida utilizada é o decibel e o valor máximo por nós suportado é de cerca de 130 dB (limiar da dor).

Altura: a altura sonora refere-se à qualidade do som no que diz respeito à sua frequência. Os sons são normalmente categorizados em graves, médios ou agudos. A altura do som é frequentemente confundida com a característica anterior, a intensidade. Quando utilizamos os termos alto e baixo referimo-nos a ondas sonoras de alta e baixa frequência, sendo assim equivalentes aos termos agudo e grave. A unidade de medida normalmente utilizada para a altura é o Hertz ou, num contexto musical, a escala das notas musicais. Genericamente, o ouvido humano pode captar sons cuja frequência esteja compreendida entre os 20 e os 20.000 Hz.

Timbre: timbre é a qualidade distintiva de um som. É a característica que confere aos sons uma “personalidade” ou “identidade” própria. Permite que a mesma nota tocada por diferentes instrumentos seja ouvida de forma distinta. Um exemplo que nos ajuda a compreender melhor o conceito de timbre é a voz humana. Duas pessoas a ler o mesmo texto são perfeitamente identificáveis. Em termos físicos, o timbre é determinado pela estrutura harmónica do som, ou seja, pela quantidade e intensidade de frequências derivadas da frequência fundamental. A tentativa de caracterização tímbrica dos sons é muitas vezes efectuada pela utilização de adjektivos como suave, estridente, cheio, incisivo ou metálico.

1.2 A Imagem

Do ponto de vista da linguagem, segundo Cloutier [10], todo o mundo visível que nos rodeia é denominado de *eidosfera*. Esta é composta por 4 componentes essenciais:

1. *Fonte luminosa*: fonte que gera e emite luz
2. *Fonte visual*: objecto que recebe e reflecte informação luminosa
3. *Luz*: meio de comunicação entre objecto e receptor
4. *Visão*: captação e processamento da luz por parte do receptor

Podemos aqui notar uma diferença fundamental da linguagem visual relativamente ao áudio. A fonte visual, contrariamente à fonte sonora, não funciona como um emissor de “energia” própria; é apenas um objecto tornado visível através da incidência da luz na sua superfície. É a reflexão desta incidência que captamos e que nos permite perceber o objecto.

Outra característica inerente ao visual é o conceito de imagem. A imagem é uma representação materializada num suporte bi ou tridimensional. O desenho numa folha de papel é uma imagem; o papel apenas possui significado em função da representação visual nele desenhada, tornando-se o desenho “o” objecto.

Contrariamente à linguagem áudio, possuidora de uma indissociável relação com o domínio temporal, a linguagem visual é essencialmente espacial. O objecto visual encontra-se sempre situado no espaço, onde usufrui das 3 dimensões, mesmo quando a profundidade é representada através da ilusão de perspectiva. A percepção visual permite a comunicação de inúmeras informações estruturadas no campo visual de forma simultânea.

O olho possui um campo visual restrito (cerca de 180° na horizontal, 150° na vertical), contrariamente ao campo auditivo, onde temos a capacidade perceber som proveniente de qualquer direcção. O campo visual é observado segundo um determinado ponto de vista. Este, aliado à estruturalidade da visão permite o estabelecimento de relações entre os vários elementos visuais constituintes de um campo visual e a focagem de atenção sobre um desses elementos em detrimento dos restantes.

A distância é um conceito essencial na estruturação da linguagem visual pois determina a relação entre o receptor e os objectos bem como a relação entre os diversos objectos de um campo visual. A distância influencia o volume relativo destes objectos graças ao efeito de perspectiva, necessário à estruturação do campo visual.

Cloutier distingue 3 parâmetros essenciais para a comunicação na *eidosfera*: a *forma*, o *tamanho* e a *matéria*.

1. *Forma*: a forma de um objecto define os seus contornos, os seus limites físicos. Ela ocupa as 3 dimensões espaciais (altura, largura e espessura) e delimita o volume de um determinado objecto.
2. *Tamanho*: o tamanho de um objecto varia consoante as dimensões da sua altura, largura e profundidade.
3. *Matéria*: a matéria diz respeito a todas as propriedades de um objecto que lhe conferem uma identidade própria. Essas propriedades poderão influenciar parâmetros como a cor, a textura, a reflexão, refração ou a transparência de um determinado objecto.

1.2.1 A visão

O olho é o órgão responsável pelo início do processo que entendemos por visão, processo esse que, no caso do ser humano, é responsável pela maior parte das informações que somos capazes de captar.

Quando se olha para um objecto, são reflectidos raios de luz desse objecto para a córnea, local onde se inicia o processo que entendemos como visão. Os raios de luz são refractados e focados pela córnea, cristalino e vítreo. A função do cristalino é a de fazer com que esses raios sejam focados de forma nítida sobre a retina. Ao atingi-la, os raios de luz são convertidos em impulsos eléctricos que, através do nervo óptico, são transmitidos para o cérebro, onde a imagem é interpretada pelo córtex cerebral [13].

Córnea e Esclera

A córnea e a esclera consistem em tecidos duros de protecção, que compõem a capa exterior do globo ocular. A esclera é a parte branca do olho. A córnea é transparente, situa-se na frente do olho, na sua parte colorida, e permite que raios de luz possam entrar no globo ocular através da pupila. Nesse globo, a esclera ocupa 85 por cento e a córnea aproximadamente 15 por cento.

Íris e pupila

A íris é o tecido que se vê atrás da córnea e é responsável pela coloração visível dos olhos. No meio da íris existe uma abertura circular, a pupila. É através da pupila que os raios de luz atingem a retina. A pupila varia de tamanho consoante a luminosidade do ambiente ficando muito pequena quando há muita luz.

Cristalino

O cristalino situa-se directamente atrás da íris e é uma estrutura flexível e transparente com o tamanho e a forma de uma aspirina. Os músculos do corpo ciliar fazem ajustes constantes na forma do cristalino. Tais ajustes servem para que a imagem se mantenha nítida sobre a retina, sempre que se muda a distância focal. A córnea e o cristalino são nutridos e lubrificados por um fluido transparente chamado humor aquoso, que enche a área entre estes.

Humor vítreo

O vítreo é uma substância translúcida com um aspecto semelhante a um gel e representa 2/3 do volume e peso do olho. Ele preenche o espaço entre o cristalino e a retina. Não tem elasticidade e é importante para manter a forma do olho sendo fundamental que se mantenha transparente para que a imagem chegue em boas condições à retina.

Retina

A retina situa-se na camada mais interna do globo ocular. É uma camada celular transparente e delicada que consiste em dez camadas, contendo três tipos de tecido: neuronal, glial, e vascular. A componente neuronal consiste nas células foto-receptoras, onde são convertidos os sinais luminosos em impulsos nervosos, posteriormente transmitidos pela camada de fibras nervosas que constituem o nervo óptico até ao córtex visual, situado na parte posterior do cérebro. A camada dos foto-receptores é composta por dois tipos de células, os bastonetes, que se distribuem principalmente pela periferia da retina e nos permitem ver em condições de baixa luminosidade dando uma vaga impressão dos objectos mas não distingui-los, e os cones, que povoam principalmente a região central da retina conhecida por mácula e nos permitem ver sob condições de alta luminosidade, sendo responsáveis pela visualização de cores e pela acuidade visual.

1.2.2 O Fenómeno Físico

O fenómeno luminoso é fisicamente mais complexo do que o fenómeno sonoro. Assim como o som, a luz possui uma natureza vibratória, mas ao contrário daquele, possui ainda uma natureza corpuscular. Quanto ao carácter vibratório, o tipo de ondas associadas à luz são as ondas electromagnéticas. No que diz respeito ao carácter corpuscular, a luz emite minúsculas partículas de energia – os fotões. A velocidade de propagação da luz é extremamente elevada (aproximadamente 300.000 Km/s). Esta propagação, contrariamente ao som, pode ser feita não só através de um meio como também no vácuo, graças à sua natureza corpuscular [13].

Existem vários modelos de representação da luz/cor, mas aquele que consideramos mais intuitivo ao nível da compreensão mas também da interacção caracteriza como principais propriedades o *brilho*, a *matiz* e a *saturação* (Figura 1).

Brilho: o brilho, também denominado de luminosidade ou apenas valor é a intensidade da luz. Está intimamente relacionada com a intensidade física da fonte luminosa. Em termos práticos, consideramos a iluminação como forte ou fraca. Quanto menor o brilho, mais escuro será o resultado (tendência para o preto); quanto maior, mais claro será (tendência para o branco).

Matiz: A matiz, também referida habitualmente como cor ou tonalidade corresponde em termos físicos à zona espectral dominante na composição do espectro electromagnético visível. Este está situado entre as ondas infra-vermelhas e ultra-violetas (700-400 milimícrones) [13].

A luz branca contém todas as cores, cada uma com um comprimento de onda distinto. Quando a superfície de um objecto é atingido por uma fonte luminosa, filtra determinados componentes espectrais da luz e reflecte os restantes. É este novo conteúdo espectral reflectido que nos dá a percepção da cor do objecto.

Saturação: a saturação diz respeito à pureza de determinada cor, ou seja, à relação entre um determinado comprimento de onda e o restante espectro electromagnético. Em termos práticos isto traduz-se na quantidade de luz branca (totalidade do espectro visível) associada à cor dominante. É através da saturação que podemos distinguir um rosa de um vermelho ou uma cor mais “viva” de uma mais “neutra”.

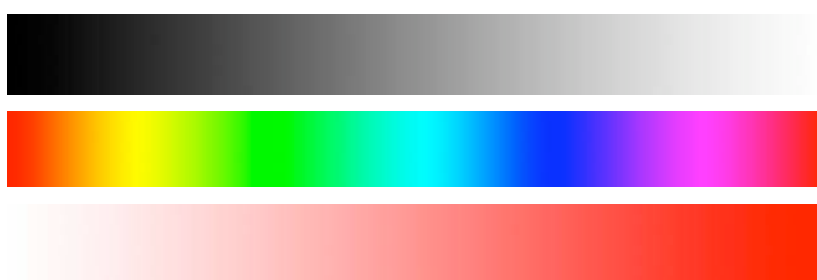


Figura 1 – De cima para baixo: variações de brilho, matiz e saturação

1.3 Estímulo e Sensação

As palavras estímulo e sensação são utilizadas para definir duas realidades distintas. Como a Tabela 1 resume, podemos verificar que o estímulo está relacionado com a fonte e com o meio de transmissão (fenómeno físico) enquanto que a sensação está intimamente relacionada com a percepção do receptor (fenómeno psicofísico).

Tabela 1 – Relação entre estímulo e sensação

Estímulo		Sensação	
Geração e propagação		Percepção	
Fenómeno físico		Fenómeno psicofísico	
Som	Luz	Audição	Visão

A informação proveniente do mundo à nossa volta é percebida graças aos nossos sentidos. Cada um dos nossos órgãos sensoriais responde aos estímulos com um determinado alcance. A percepção não envolve apenas a recepção da informação pelos órgãos sensoriais, mas também a codificação, transmissão e sobretudo o processamento dessa informação pelo sistema nervoso central. Sabemos que a forma como

sentimos os estímulos não é linear. G.T. Fechner passou muitos anos a tentar determinar relações quantitativas entre o estímulo e a sensação percebida, tendo resumido o seu trabalho numa simples relação matemática a que hoje chamamos *Lei de Fechner*: à medida que o estímulo aumenta por multiplicação, a sensação aumenta por adição [12]. Como exemplo, à medida que a intensidade de uma fonte sonora vai duplicando, a percepção que daí resulta é apenas a do aumento de um valor na escala (Tabela 2). Utilizando uma terminologia mais matemática, trata-se de uma relação logarítmica; a lei de Fechner determina que a sensação cresce com o logaritmo do estímulo. Esta é a principal razão para a utilização do *Decibel* na medição acústica.

Tabela 2 - Exemplos de valores da intensidade sonora quantificada num escala linear e logarítmica

Escala linear	10^{-8}	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}	0,01	0,1	1,0	10
Escala logarítmica (dB)	40	50	60	70	80	90	100	110	120	130

No decorrer desta tese, esta diferenciação não será amplamente abordada de forma explícita, mas possuirá um papel importante na compreensão de vários fenómenos auditivos e visuais, compreensão essa fundamental para o relacionamento de características entre estes dois sentidos. Ao considerarmos esta informação, os relacionamentos efectuados poderão ser mais ajustados, já que a natureza da escala de determinado parâmetro poderá ser tida em conta.

1.4 Correspondências áudiovisuais?

No ano de 1690, John Locke, um filósofo inglês, reportou no seu livro intitulado *An Essay Concerning Human Understanding* que um homem cego lhe teria afirmado que o som do trompete lhe aparentava a cor escarlate. Esta afirmação motivou uma ampla discussão acerca da possibilidade de existirem correspondências entre o fenómeno sonoro e o luminoso [14]. Uma das pessoas a explorar esta possibilidade foi o famoso poeta e filósofo, Goethe (1749-1832), apoiado na crença comum de que tanto o som como a luz eram o resultado de vibrações de um determinado meio¹ [15].

Visto que ambos os fenómenos possuem uma natureza vibratória, alguns estudiosos consideravam que se o Dó central pudesse ser transposto quarenta oitavas, seria possível visualizar a cor vermelha. Apesar desta e de outras ideias serem mais tarde completamente refutadas pela ciência, a tentativa de relacionar o que vemos com o que ouvimos (e vice-versa) parece ser algo que nos é inato. Conforme tivemos a oportunidade de afirmar anteriormente, consideramos que nenhum mapeamento de características elementares como as que serão aqui referidas será suficiente para relacionar o som e a imagem de forma efectiva. Este facto, na nossa opinião, constitui uma mais valia para os artistas, pois será necessária a combinação criativa destes

¹ Segundo Goethe, o ar, no caso do som, e o éter luminífero (*ether luminiferous*), no caso da luz.

elementos de forma a obter um produto artístico satisfatório, correlacionado nos domínios sonoro e visual. Por este motivo, a validade das relações aqui apresentadas não é por nós posta em causa. Somos da opinião que, independentemente da falta de validação científica, a combinação criativa de características sonoras e visuais tendo como propósito a expressão artística constitui uma área de exploração potencialmente bastante apelativa.

1.4.1 A cor e as notas musicais

De entre todas as correspondências feitas ao longo da história entre parâmetros sonoros e visuais, a mais habitual tem sido a de fazer corresponder a cor à altura sonora. Fred Collopy sistematiza de forma bastante eficaz estas relações efectuadas por algumas das principais personalidades que exploraram este assunto nos últimos 3 séculos (Figura 2) [16]. Enquanto que algumas destas propostas não passaram do plano teórico, outras tiveram consequência prática, tendo sido materializadas em instrumentos pelos seus autores.

Em 1912, Wallace Rimington, no seu livro *Colour-Music: The Art of Mobile Colour* [17] afirma que mesmo que se possa demonstrar inequivocamente que não existe qualquer relação entre as escalas de cor e de sons musicais, a validade da *Colour-Music* enquanto forma de arte separada e distinta e modo de pesquisa experimental permaneceria inafectada. Considera ainda que “a pergunta acerca de uma possível analogia entre as duas escalas é interessante, mas até que ponto se sustenta não é de grande importância.”

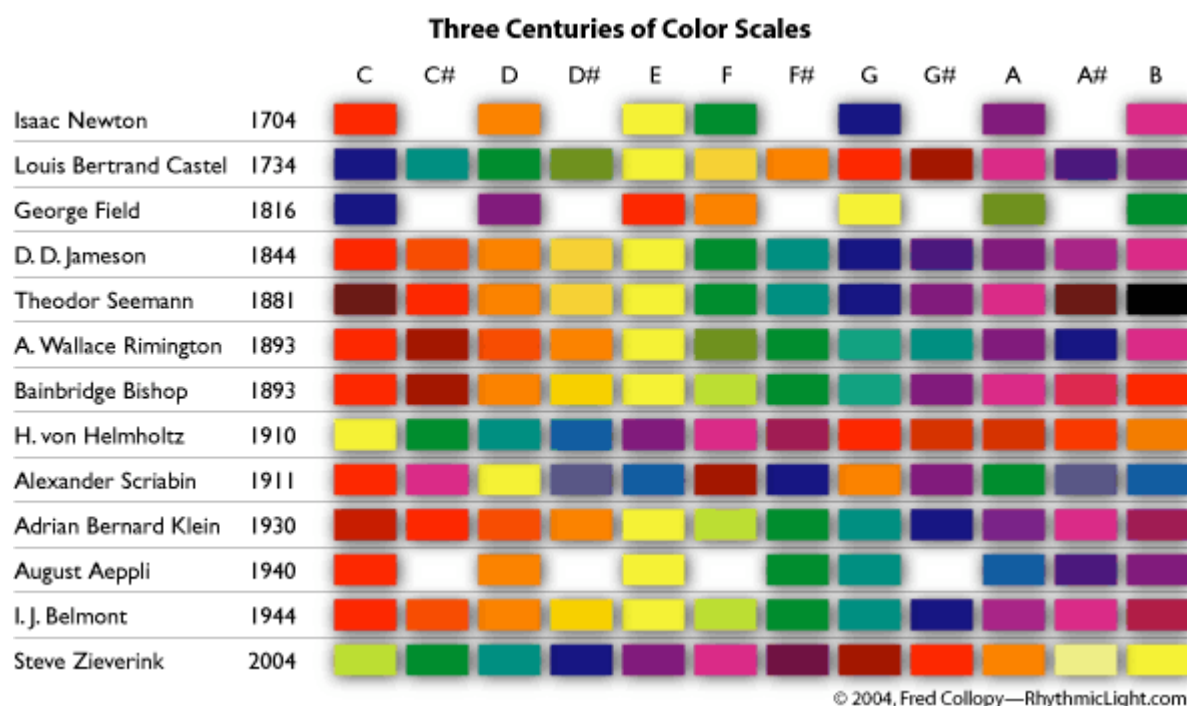


Figura 2 – Quadro comparativo das associações efectuadas entre altura musical e cor ¹

¹ Imagem de Fred Collopy retirada do URL: <http://www.rhythmiclight.com/archives/ideas/colorScales.gif>

1.4.2 A cor e o timbre

Kandinsky, no seu livro *On the Spiritual in Art* (1912) [18], contribuiu para uma mais rica possibilidade de relacionamentos entre som e cor ao relacionar a cor com o timbre sonoro de determinados instrumentos. Na Tabela 3 podemos ver as principais associações por ele sugeridas. No entanto, Kandinsky não foi o único a sugerir o relacionamento entre estes dois parâmetros. Douglas Jones, no seu livro *The Art of Light & Color* (1972) refere que Joachim Raff e Christopher Ward expressaram também as suas opiniões relativamente a este assunto [19]. Dado o carácter extremamente subjectivo e pessoal deste tipo de mapeamentos (uma vez que não são suportados por qualquer semelhança em termos de fenómeno físico), as abordagens de Kandinsky, Raff e Ward são, previsivelmente, distintas e às vezes até contraditórias.

Tabela 3 – Relações de Kandinsky entre a cor e o timbre sonoro

Cor	Som
Amarelo	Trompete a tocar forte
Laranja	Sino de igreja, voz contralto, viola
Vermelho	Fanfarras, tuba, ritmo de tambores
Púrpura	Sons agudos do violino, pequenos sinos
Violeta	Corne inglês, charamela, fagote
Azul	Flauta, violoncelo, contrabaixo, órgão baixo
Verde	Sons calmos e meditativos do violino

Apesar da impossibilidade de atribuição de uma relação exacta entre estes parâmetros, consideramos que a combinação destes pode ter algum fundamento. É que, ao contrário da altura das notas, o timbre sonoro, como a cor, é uma característica facilmente identificável. Exceptuando os músicos com o chamado “ouvido absoluto”, a vasta maioria não é capaz de identificar a altura de determinada nota quando esta se encontra isolada de um determinado contexto. Sem qualquer referência, somos normalmente incapazes de identificar se determinada nota é um Dó ou um Ré. Pelo contrário, o timbre de um instrumento ou de uma voz, é algo facilmente perceptível por todos. Parece-nos por este motivo que, apesar da escolha ser extremamente pessoal e até eventualmente arbitrária, o relacionamento entre timbre sonoro e cor poderá fazer algum sentido em termos de identidade.

1.4.3 A cor e o modo

Frank Popper, em *Origins and Development of Kinetic Art* (1968), descrevendo as experiências realizadas por Scriabin nesta área provavelmente no contexto da sua obra *Prometheus*, refere que este defendia que “a cada modo correspondia uma matiz de cor específica, e a cada modulação uma variação desta matiz. Alterações de um modo maior para um menor poderiam portanto ser sublinhadas por fortes contrastes, tanto a um nível visual como cromático.” [20]

D.S. Myers, um psicólogo que costumava falar com Scriabin disse certa vez que este compositor, acompanhado pelo seu amigo Rimsky-Korsakoff num concerto em Paris, comentou que a música que ouviam (na tonalidade de Ré Maior) lhe lembrava a cor amarela. Korsakoff respondeu que, para ele, fazia-lhe pensar na cor dourada. Desde essa altura, Scriabin passou a comparar com outros músicos esta relação entre cor e tonalidade. Scriabin estava convicto de que o relacionamento por ele percebido entre música e cor deveria ser similar em todos aqueles que possuíssem o mesmo dom sinestésico.

1.4.4 Associações a outras características da cor

Apesar das associações mais recorrentes entre o som e a cor terem sido já mencionadas, estas centraram-se maioritariamente na característica da cor que designamos por matiz. Existem contudo possibilidades de relacionamento que podem ser também efectuadas com outras características da cor como o brilho ou a saturação. Como exemplo, Karl Gestner, no seu livro *The Forms of Color* (1986) relacionou a amplitude sonora com a saturação da cor. Equiparou as cores puras (elevado grau de saturação) aos sons de grande amplitude e as cores com uma tonalidade mais neutra a sons de pequena amplitude [21]. Comparou ainda a altura sonora com a intensidade da luz, ao referir que “*estamos habituados a ver as cores escuras como sons graves, as claras como agudos.*”

1.4.5 O movimento e a representação temporal

Apesar de pintores e escultores terem tentado sugerir movimento nas suas obras estáticas, só com o aparecimento do cinema é que o puderam concretizar. A representação temporal no espaço visual é-nos, por motivos culturais, muito próxima. Na cultura ocidental, a leitura de informação, seja ela textual ou até musical, é feita no sentido da esquerda para a direita e, posteriormente de cima para baixo. No contexto de artes visuais como a pintura, dada a sua natureza estática, este assunto ganha particular relevância, uma vez que o movimento tem necessariamente de ser sugerido. Com a utilização de outras técnicas como o cinema e o vídeo, que compreendem já o movimento, a representação temporal no domínio espacial pode continuar a ter alguma importância uma vez que nos permite obter uma visão geral, simultânea, de tudo o que se passa num determinado intervalo de tempo.

Segundo Fred Collopy [2], “*é apenas através da memória que nós experienciamos o movimento, e para construir uma experiência coerente no tempo, o artista tem de auxiliar o espectador a dar significado às várias imagens sucessivas.*” Esta afirmação parece apoiar o facto de ser por vezes necessário sublinhar o desenvolvimento temporal em termos espaciais, através da persistência visual de eventos passados ou de outros recursos. Collopy reafirma esta opinião citando Rudolf Arnheim [22], que diz: “*Em cada momento específico, poderemos não saber o que virá a seguir, mas não podemos demitir da nossa consciência o que ouvimos ou vimos anteriormente.*”

Gyorgy Kepes, no seu livro *Language of Vision* [23], considera que a música se assemelha mais a uma pintura ou a um desenho de arquitectura do que qualquer outro tipo de manifestação artística. Segundo ele, *“a diferença é que num desenho as linhas são visíveis e constantes, enquanto que na música elas são audíveis e em movimento. As notas isoladas são os pontos através dos quais são desenhadas as linhas; e a impressão pretendida, e apreendida pelo ouvinte inteligente, não é a de notas isoladas, mas a de linhas contínuas de notas, descrevendo movimentos, curvas e ângulos, subindo, descendo – directamente análoga às impressões lineares reveladas por uma pintura ou desenho.”*

1.4.6 O som e a forma

A forma dos elementos visuais pode estar relacionada com alguma característica sonora. Franz Boas, no seu livro de *General Anthropology* (1938) é da opinião de que os sons agudos aparentam ter um tamanho reduzido, enquanto que os sons graves sugerem um tamanho maior, opinião que aparenta ser generalizada.

Kandinsky defendia em *Point and Line to Plane* (1926) que para além da cor, a forma pode também estar relacionada com a música. Considerava que a espessura de determinada linha poderia estar relacionada com a tessitura de determinado instrumento. Assim, *“o violino, a flauta e o piccolo produzem uma linha fina, a viola e o clarinete uma linha sensivelmente mais espessa; e através dos instrumentos mais graves, chegamos a linhas cada vez mais espessas, até chegar às notas mais graves de um contrabaixo ou de uma tuba.”* Identificou o órgão como um instrumento de linha e o piano como um instrumento de ponto.

Defendia ainda que *“a escala de valores do pianissimo ao fortissimo podem ser expressas através do aumento ou diminuição da intensidade da linha, ou pelo seu grau de claridade. A pressão da mão no arco corresponde de forma perfeita à pressão da mão no lápis.”* [24]

Lawrence E. Marks refere em *The Unity of the Senses* (1978) que Karwoski e Odbert descobriram em 1938 uma relação sistemática entre a forma visual e o andamento da música. Segundo eles, quanto mais rápida for a música, mais “afiada” e mais angular será a imagem resultante [25].

1.4.7 Um nível superior

Até agora, têm sido abordados alguns dos mapeamentos mais recorrentes. Tendemos a considerar a maior parte destes como mapeamentos básicos, devido ao carácter elementar dos parâmetros envolvidos: altura, intensidade e timbre sonoro, matiz, saturação, intensidade da luz. Incapazes de aceitar a validade deste tipo de mapeamento entre o fenómeno sonoro e o visual, muitas personalidades que se debruçaram sobre o assunto crêem que apenas através da adopção de um nível superior de relacionamento se torna possível perceber a união desejada entre estes dois sentidos. Tal é o caso de Lawrence E. Marks, ao afirmar que sendo a música uma forma de arte complexa, não será de estranhar que as imagens visuais por ela sugeridas

se possam relacionar com múltiplas características musicais [25]. Segundo o mesmo, as respostas visuais à música podem variar de acordo com a tonalidade, padrão ou progressões musicais.

Acerca da relação entre cor e melodia, Goethe refere que dadas as frequentes comparações com que nos deparamos, “*por vezes como alusões passageiras, outras como paralelismos circunstanciais*”, se prova a existência de uma certa relação entre as duas [15].

Maitland Graves, em *The Art of Color and Design* [26], considera que apesar da luz e do som possuírem algumas semelhanças, a formulação de um sistema de *Colour Music* baseado na escala musical não funciona. Graves cria que este tipo de abordagens peca por não ter em conta as diferenças fundamentais entre a audição e a visão, não refutando contudo a possibilidade de uma arte visual semelhante à música se “*fosse baseada nos mesmos princípios estruturais que a música, mas desenvolvida de acordo com as leis que governam o olho e a luz.*”

1.4.8 Opiniões contrárias

Apesar das opiniões atrás manifestadas, que defendem a existência de alguma relação entre o fenómeno sonoro e o luminoso, existem também outras opiniões divergentes que, de forma igualmente válida, refutam a existência desse relacionamento. Estes afirmam que o som e a luz são de tal forma diferentes, que são dois fenómenos tão distintos, que qualquer teoria de correspondências deveria ser abandonada. É importante referir que esta refutação de correspondências situa-se na maior parte dos casos apenas no plano científico, não negando os autores explicitamente a validade das mesmas num contexto de exploração artística.

Frank Popper, no seu livro *Origins and Development of Kinetic Art* [20], refere que Wilfred, ao deparar-se com as divergências na forma como Newton e Castel relacionaram a cor com o som, ao conhecer a rejeição de Goethe relativamente a uma comparação directa entre cor e som, e ao analisar o relativo insucesso dos instrumentos de Bishop, Rimington e Scriabin, optou por concentrar-se no desenvolvimento de “*uma arte da luz, onde som e música fossem completamente excluídos ou incluídos como meros acessórios.*” Comparando as notas musicais com as cores, Karl Gerstner, em *The Forms of Color* [21] afirma que “*as cores possuem um maior grau de individualidade, ou personalidade, que as notas musicais.*” Ogden N. Rood, no seu livro *Modern Chromatics* [27], já em 1879 enuncia algumas diferenças fundamentais entre estes dois fenómenos:

- a sensação sonora está particularmente ligada com o tempo, a da visão com o espaço.
- a nossa percepção da cor não se estende sequer acima de uma oitava, enquanto que na música, são utilizadas 7 oitavas.
- quando ouvimos dois sons musicais, sentimos consonância ou dissonância, e o ouvido consegue reconhecer separadamente as notas tocadas; mas quando duas “massas” de luz se intersectam, é produzida uma nova cor, na qual os constituintes originais não se conseguem distinguir.

- na música, os intervalos são relações definitivas e facilmente reconhecíveis, como por exemplo a da fundamental com a sua quinta ou oitava; podemos calcular idênticas relações para as frequências da luz, mas as cores nunca seriam identificadas de forma precisa.
- na pintura, é comum a utilização de gradações de cor da forma mais progressiva possível, mas um procedimento semelhante na música daria lugar a sonoridades indesejáveis, pois as notas não fariam parte da escala musical estabelecida.

Segundo o autor, *“estes factos, que estão susceptíveis à prova mais sólida, poderão bastar para mostrar que existe uma diferença fundamental entre as sensações da visão e da audição, e que qualquer teoria da cor baseada na nossa experiência musical deverá assentar em fantasia em vez de facto.”*

Roger Scruton, no seu livro *The Aesthetics of Music* [28] refere que apesar de, do ponto de vista do fenómeno vibratório, a altura de um som poder ser comparável à cor de uma luz, tanto uma como outra possuem um conjunto de características distintas:

- Percebemos a variação de altura como se de uma dimensão se tratasse, efectuando movimentos ascendentes e descendentes num espaço bi-dimensional.
- As várias alturas das notas são agrupadas apenas por uma relação de proximidade. As distâncias entre elas são a única forma de as compararmos enquanto notas. Não existem grupos de notas que se considerem possuidoras de uma determinada identidade, contrariamente às cores, onde, de forma natural tendemos a identificá-las, separá-las e agrupá-las como fazendo parte dos vermelhos, laranjas, amarelos, verdes, etc.
- Finalmente, o espectro audível possui a experiência particular da repetição, ou seja, da oitava. Ao chegar à oitava, voltamos ao ponto de partida. Este fenómeno será menos notório em sons que percorrem de forma progressiva o espectro audível, como o *glissando* de uma sirene, do que quando avançando em pequenos passos, num contexto musical. Aliás, é assente nesta ideia de “equivalência da oitava” que todo o sistema tonal da música ocidental se baseia.

Adrian Bernard Klein opõe-se a este assunto de uma forma particularmente vincada ao recusar a *“...adesão àquele grupo (se tal existe) que suporta a doutrina de correspondências, determinadas através de análise científica, entre cores e sons. Analogias! Ah! Poderão ser analogias. Existe uma analogia entre a tabela periódica de Mendeleeff e a escala diatónica. Contudo, não observo que a nota Dó expresse uma impressão do Hélio, ou que a nota Dó uma oitava acima, seja distintamente remanescente do néon.”* [3] Também Goethe é da opinião que cor e som não podem ser comparados de forma alguma. Em *Theory of Colours* [15] refere que estes *“São como dois rios que nascem na mesma montanha, mas que a partir daí percorrem os seus cursos em condições totalmente diferentes, em duas regiões totalmente diferentes, de forma a que ao longo de todo o percurso de ambos não existem dois pontos que possam ser comparados. Ambos são efeitos gerais elementares agindo de acordo com a lei geral da separação e tendência para a união, da ondulação e oscilação, contudo agindo em províncias totalmente diferentes, de modos diferentes, em meios elementares diferentes, para sentidos diferentes.”*

1.5 Instrumentos audiovisuais

Neste espaço pretendemos referir alguns dos mais importantes instrumentos construídos com um intuito de *performance* audiovisual, similar à abordagem por nós proposta nesta tese. Na maior parte dos momentos históricos, tem existido um atraso mais ou menos significativo entre a disponibilização de novas tecnologias e a adopção das mesmas por parte dos artistas, utilizadas como ferramentas para a produção artística. No entanto, no caso concreto da *Visual Music*, nota-se que em muitas situações foi a própria necessidade de expressão artística que motivou o progresso tecnológico. A distinção entre artista, cientista ou técnico é, em muitos dos nomes de relevo desta área, impossível de ser feita. Os próprios artistas, para materializar as suas ideias, foram obrigados a construir os “instrumentos” que possibilitassem a comunicação das mesmas. No decorrer da elaboração desta tese de mestrado, sentimos igualmente a necessidade de gerir essas mesmas funções, uma vez que a construção do ponto de vista técnico/científico do sistema audiovisual proposto contem uma finalidade claramente artística, que merece ser destacada.

1.5.1 Século XVIII

O primeiro dispositivo de que temos conhecimento construído com o intuito de visualizar música é da autoria de Louis-Bertrand Castel (1688-1757), padre jesuíta, filósofo e matemático. Visto que não se considerava técnico, a sua intenção inicial era apenas a de teorizar sobre a viabilidade da construção de instrumentos de som e luz. No entanto, uma vez que as suas ideias foram encaradas com algum cepticismo, viu-se “obrigado” a construir um modelo que as exemplificasse, resultando no aparecimento do *Clavecin Oculaire* (cravo ocular) em 1734 [14]. Este instrumento combinava um tradicional cravo com a visualização de fitas de papel semi-transparentes com cores que, segundo Castel, corresponderiam às notas da escala musical ocidental. As motivações de Castel são bem resumidas na seguinte afirmação retirada do seu tratado intitulado *Optique des couleurs* (1740), também conhecido como o “tratado da melodia das cores”:

“Que aventura mais estranha poderia ser imaginada em todo o campo artístico do que tornar o som visível, tornar acessível aos olhos os muitos prazeres que a Música proporciona aos ouvidos?”

O trabalho de Castel serviu de inspiração para muitos outros compositores e artistas que, ao longo do tempo se foram interessando pela visualização da música. É interessante notar que desde esta altura o teclado tem sido o *interface* privilegiado para o controlo dos sistemas audiovisuais, eventualmente dada a sua natureza mecânica.

1.5.2 Século XIX

Durante o século XIX, existiram vários artistas/técnicos que continuaram a explorar esta forma de expressão artística. Foram concebidos vários instrumentos, entre os quais o de D.D. Jameson em 1844, provavelmente inspirado pelas sugestões de Erasmus Darwin em 1789 (avô do conhecido Charles Darwin). Peacock [14] destaca também o *Pyrophone* do inglês Frederick Kastner em 1873, ou o instrumento do americano Bainbridge Bishop, em 1877, mas considera que o instrumento mais conhecido do século XIX é muito provavelmente o *Colour-Organ* (orgão de cor) de Alexander Wallace Rimington, patenteado em 1893. No seu livro *Colour-Music: The Art of Mobile Colour* [17], Rimington revela ser da opinião de que existem analogias físicas entre os fenómenos sonoro e luminoso, e afirma que ambos “*são resultado de vibrações que estimulam o nervo óptico e aural, respectivamente.*”

Apesar destas analogias físicas, é interessante notar que Rimington relacionou a totalidade do espectro visível não com a totalidade do espectro audível, mas restrita à oitava musical, repetindo assim o ciclo de cores em cada uma das oitavas. Utilizou ainda a saturação para diferenciar as várias oitavas, sendo que quanto mais alta a oitava, mais luz branca era adicionada à cor base da nota [17]. É de salientar ainda que no instrumento de Rimington era possível configurá-lo de forma a não efectuar o mapeamento do espectro visível por oitavas, mas pela totalidade do teclado. Esta flexibilidade de mapeamento é com certeza um reflexo da sua abertura às várias possibilidades de relacionamento entre parâmetros sonoros e visuais, não sendo por isso defensor da existência de uma verdade absoluta neste domínio.

1.5.3 Século XX

O compositor russo Alexander Scriabin foi o primeiro compositor a incluir uma parte específica na sua partitura de orquestra reservada para a projecção de luz. Esta dizia respeito à sua conhecida sinfonia de 1911 intitulada *Prometheus, the Poem of Fire*. Foi desenvolvido um instrumento de projecção de luz, o *Chromala*, propositadamente para a *performance* desta obra. Consta-se que, devido a alguns problemas técnicos, bem como à pequena tela de projecção, a estreia desta sinfonia acompanhada pela projecção luminosa, em 1915, não tenha sido muito bem aceite por parte do público. No entanto, em apresentações posteriores, já com melhores condições, a crítica terá sido mais favorável [14]. Só nos inícios do século XX é que esta temática de relacionamento entre som e luz começou a ser mais explorada, devido à curiosidade nestes assuntos por parte de artistas como pintores e músicos. Estes estudavam as formas como a pintura e a música poderiam ser relacionadas. Enquanto que os criadores dos instrumentos abordados tinham o intuito de demonstrar a existência de associações directas do ponto de vista físico entre os fenómeno sonoro e luminoso, a grande maioria dos instrumentos do início do século XX não foram construídos com esse propósito. Segundo Peacock [14], “*No início da década de 20, ficou claro que não existia nenhum padrão de correspondências*

entre cores e som que fosse inquestionável. Por esse motivo, muitos previram a evolução de uma nova e independente forma de arte – a pura manipulação da luz, sem qualquer ligação com o som.”

Esta foi uma altura de incrível expansão no desenvolvimento de sistemas abstractos de *performance* visual. Com os imensos avanços em áreas tão fulcrais como a electricidade, a óptica, o cinema, e o conhecimento mais aprofundado ao nível da percepção sensorial, muitos instrumentos foram desenvolvidos. De todos, será conveniente destacar, pela sua importância, três deles: o *Clavilux*, de Thomas Wilfred, o *Lumigraph* de Oskar Fischinger, e o projector *MobilColor* de Charles Dockum.

Clavilux

Thomas Wilfred (1889-1968) começou por procurar o mapeamento absoluto entre o som e a cor, mas ao investigar e analisar o trabalho dos seus antecessores nesta área de estudo, acabou por concluir que essa correspondência exacta não existe [29]. Ao rejeitar completamente qualquer teoria que relacionasse o som com a luz, voltou então a sua atenção para o desenvolvimento de sistemas estritamente visuais, onde som e música poderiam funcionar apenas como acessórios [14]. Desenvolveu um instrumento denominado de *Clavilux* em 1919 que consistia mais uma vez num teclado, desta vez de órgão, no qual eram acrescentados 6 projectores e alguns reflectores adicionais. É interessante referir que uma versão reduzida deste instrumento foi desenvolvida, o *Home Clavilux*. Este instrumento assemelhava-se a uma tradicional televisão e, como o nome indica, foi concebido para uma utilização doméstica.

A concentração de Wilfred na componente visual da chamada *Colour-Music* permitiu elevá-la a uma forma de arte autónoma a que o próprio designou de *Lumia*, em 1905, e cujo objectivo era o de permitir aos artistas visuais manipular as imagens da mesma forma que os músicos manipulam os sons, ou seja, “música para os olhos”. Assim como a música, uma obra visual criada pode possuir uma estrutura composicional pré-definida ou ser essencialmente improvisada. Wilfred começou por utilizar a cor de uma forma completamente abstracta, mas rapidamente chegou à conclusão de que a forma e o ritmo eram elementos igualmente essenciais para a criação de um resultado visual apelativo. Assim, segundo ele, a concepção de instrumentos visuais deve ter como principais elementos de referência a cor, a forma e o movimento [29]. A utilização destas três dimensões tem ainda como vantagem permitir a aplicação de todo o conhecimento teórico acumulado até aos dias de hoje no domínio das artes visuais, pois estes são considerados elementos estruturantes fundamentais. Muito provavelmente, devido à relevância que Wilfred atribuiu a estas novas dimensões, forma e ritmo, a opinião pública foi, mais do que nunca, favorável às inúmeras exibições por ele realizadas por toda a Europa, Estados Unidos e Canadá. Todos ficaram de tal forma impressionados, que os próprios críticos tiveram dificuldade em descrever o resultado da actuação de Wilfred no seu *Clavilux*. Como exemplo, Deems Taylor referiu que “*É difícil de expressar em palavras. Descrever o Clavilux a alguém que não o viu é como descrever uma laranja a um esquimó.*” [14]

É importante referir que apesar da maioria dos concertos de Wilfred se realizarem em silêncio, ele participou também em concertos onde colaborou com compositores e músicos. Um exemplo disso é a sua colaboração na obra *Scheherazade* de Rimsky-Korsakov, em 1926.

Lumigraph

Oskar Fischinger (1900-1967), um nome incontornável nesta área, concebeu no final dos anos 40 um órgão de cor (*color organ*) [4]. Ao contrário de Thomas Wilfred, que ao concluir não haver uma correspondência absoluta entre som e imagem desistiu da procura de relacionamentos entre estes, Fischinger era da opinião que o mapeamento de correspondências entre estes dois meios possui um carácter profundamente subjectivo e pessoal. Por isso, para ele, o próprio mapeamento entre parâmetros sonoros e visuais revelava-se já à partida como uma oportunidade para a expressão artística [30], opinião aliás, com a qual concordamos inteiramente e que funciona como premissa fundamental para a própria tese.

MobilColor

Este “instrumento” foi concebido por Charles Dockum, um construtor de órgãos de cor no final dos anos 30. O *MobilColor* consistia num sistema de projecção que permitia a criação de padrões temporais de formas coloridas. A organização temporal destes padrões era definida através de um método mecânico de programação que proporcionava diferentes formas [30]. Estas formas estavam já previamente concebidas, logo as possibilidades de variação visual das *performances* com este instrumento estavam de certa forma limitadas à alterações destas formas através de efeitos de rotação, translação, escalamento, colorização e desfocagem. Podemos de certa forma comparar este método de produção visual com a técnica de *sampling* tão utilizada actualmente, onde os elementos básicos de geração (*fonte*) são também previamente gravados, sofrendo alterações mediante a acção de processamento adicional (*modificadores*).

1.5.4 Presente

Nos últimos anos, os computadores, com as suas imensas capacidades ao nível do processamento, da visualização e sonorização tornaram possível, mais do que em qualquer outro momento da história, a criação de instrumentos audiovisuais, proporcionado assim um certo “revivalismo” nesta área da visualização da música. Quer ao nível do entretenimento, quer a um nível mais artístico/cultural, os consumidores estão cada vez mais habituados a usufruir de conteúdo multimédia, no qual som e luz, ao estar relacionados com os principais órgãos sensoriais, possuem um papel fundamental. Quer ao usufruir dos efeitos visuais gerados pelas comuns *jukeboxes* musicais como o *iTunes*, *Winamp* ou *MediaPlayer*, ao assistir a um grande espectáculo de *rock* com programações de luzes, *lasers* e *videowalls*, ao ver um *videoclip* ou ao visitar uma instalação multimédia exposta num museu, a fruição audiovisual é algo a que estamos constantemente

sujeitos. Enquanto que anteriormente os mecanismos de geração e processamento dos instrumentos audiovisuais possuíam um suporte físico, actualmente estes dependem cada vez mais de componentes de *software*. Seja através de linguagens de programação tradicionais (C, C++, Java, etc), ambientes gráficos de programação (Max/MSP¹, Pd², etc.) ou aplicações dedicadas (Arkaos GrandVJ³, Modul8⁴, VDMX⁵, etc.), é hoje possível aos artistas, mais do que nunca, expressarem-se sem grandes impedimentos em termos de limitações tecnológicas ou de custos financeiros.

Segundo Peacock [14], “Cada geração, parece, tem que redescobrir e redefinir a arte da Colour-Music por si própria.” Embora concordemos com esta afirmação, somos também da opinião que, apesar das tecnologias poderem divergir, a experiência acumulada por todos os músicos, artistas, inventores e cientistas que exploraram esta área constitui certamente uma mais valia para todos aqueles que pretendam aprofundar e desenvolver esta mesma área. A Tabela 4 menciona, em forma de conclusão, alguns dos mais importantes instrumentos para a *performance* audiovisual. Entre os instrumentos referidos, é importante salientar o facto de uns produzirem som e imagem (luz) enquanto que em outros instrumentos, os seus construtores concentraram-se apenas na construção da componente visual. Neste último caso, a música continua a ter um papel essencial, servindo como fonte de inspiração para a *performance* visual, que por sua vez, acaba por complementar a música.

Tabela 4 – Alguns dos mais importantes instrumentos para *performance* audiovisual

Ano	Nome	Inventor	Som	Luz
1734	<i>Clavecin Oculaire</i>	Louis-Bertrand Castel	✓	✓
1869	<i>Pyrophone, Singing Lamp</i>	Frederic Kastner	✓	✓
1877	Orgão de cor	Bainbridge Bishop	✗	✓
1893	<i>Color Organ</i>	Alexander Wallace Rimington	✗	✓
1919	<i>Sarabet</i>	Mary Hallock-Greenewalt	✓	✓
1919	<i>Clavilux</i>	Thomas Wilfred	✗	✓
1920	<i>Piano Optophonique</i>	Wladimir Baranoff-Rossiné	✗	✓
1925	<i>Sonchromatoscope</i>	Sandor Alexander Lazlo	✓	✓
1928	Teclado de Cor	Zdenek Pesanek	✗	✓
1933	Projector de Cores	Adrian-Bernard Klein	✗	✓
1940	Projector <i>MobilColor</i>	Charles Dockum	✗	✓
1950	<i>Lumigraph</i>	Oskar Fischinger	✗	✓
1960	<i>Sonovision</i>	LLoyd G. Cross	✓	✓
1966	<i>Chromara</i>	Richard Land	✓	✓
1974	<i>Vampire</i>	Laurie Spiegel	✓	✓

¹ <http://www.cycling74.com/products/maxmsp>

² <http://puredata.info/>

³ <http://www.arkaos.net/>

⁴ <http://www.modul8.ch/>

⁵ <http://www.vidvox.net/>

1.6 A Música e as Artes Visuais

A música tem tido ao longo do tempo uma grande importância nas artes visuais. Por este motivo, aproveitamos este espaço para referir alguns artistas do domínio visual que se deixaram influenciar por ela como fonte de inspiração artística para as suas obras, quer na pintura, quer no cinema e vídeo.

Como referimos anteriormente, nos inícios do século XX, muitos artistas acreditavam que a música era a mais livre e mais libertadora forma de expressão criativa. Esta opinião não era contudo sentida apenas no domínio das artes visuais, como podemos comprovar pela opinião de Oscar Wilde, o qual afirmou que *“a música é a arte em que forma e matéria são sempre uma, a arte em que o assunto não pode ser separado do método da sua expressão, a arte que de modo mais completo concretiza o ideal artístico, é a condição à qual todas as outras artes permanentemente aspiram.”* [31] Reforçando esta ideia, também Walter Pater afirmou em 1877 que *“toda a arte aspira constantemente em direcção à condição da música.”* [32]

Ao contrário dos instrumentos audiovisuais atrás referidos, onde a *performance* em tempo-real estava subjacente à sua utilização, a pintura e o cinema não possuem essa necessidade, na medida em que um quadro ou um filme podem ser idealizados e concebidos sem quaisquer restrições temporais e de forma autónoma relativamente à música. Embora a abordagem desta tese incida particularmente sobre as implicações da visualização do som em tempo-real, pela importância histórica e pelo contributo que estas formas de expressão artística tiveram na definição de novas estéticas visuais, mais tarde aproveitadas por instrumentos audiovisuais, consideramos ser de grande relevância para esta tese uma breve contextualização a este importante assunto.

1.6.1 Pintura

Os pintores, ao desejar criar uma estética nas artes visuais similar à música, começaram a criar pinturas abstractas, não representativas, cheias de formas e cores vibrantes, com sugestões de movimento e ritmo. Um exemplo disso é a pintura de 1913 intitulada *“Capriccio Musicale”* (Figura 3), na qual o pintor russo, escultor e músico Baranoff-Rossiné tenta captar de forma visual a essência da música de circo [33]. Para além do ritmo provocado pelas cores e formas utilizadas, podemos observar ainda a utilização de excertos da partitura da *Rapsódia Hungária Nº2*.



Figura 3 - “Capriccio Musicale” (1913) de Baranoff-Rossiné

Um outro pintor russo, Wassily Kandinsky (Figura 4), um dos primeiros criadores da abstracção pura na pintura moderna, era também músico. Este facto faz-se notar claramente nas suas pinturas, que reflectem uma forte inspiração na linguagem musical. Esta influência é até notória na escolha dos títulos para os seus quadros, onde utiliza termos musicais como “Improvisações”, “Impressões” ou “Composições.” Kandinsky utilizava a cor de uma forma bastante relacionada com o fenómeno sonoro, pois associava a cor à altura (afinação), a tonalidade ao timbre sonoro, e a saturação ao volume sonoro. É muito conhecida a sua afirmação de que *“quando via cor, ouvia música.”*

“A cor influencia directamente a alma. A cor é o teclado, os olhos são os martelos, a alma é o piano com muitas cordas. O artista é a mão que toca uma ou outra tecla, para causar vibrações na alma.”

“O preto é como o silêncio do corpo após a morte, o fecho da vida.”

Wassily Kandinsky, 1912 [18]



Figura 4 – “Contrasting Sounds” (1923) de Wassily Kandinsky

Paul Klee é um outro exemplo de um pintor e músico que tentou utilizar elementos expressivos próprios da linguagem musical nas suas pinturas. A pintura *“Fugue in Red”*, de 1921 (Figura 5), representa claramente uma forma musical, com o tempo representado no eixo horizontal, da esquerda para a direita, tal como o sentido de leitura de uma partitura musical.



Figura 5 – “Fugue in red” (1921) de Paul Klee

Para além dos pintores referidos, existem ainda muitos outros como Mondrian, Kupka, Gleizes ou Hans Richter que, de uma forma ou de outra, exploraram algumas ligações possíveis entre a pintura e a música. No entanto, à medida que a pintura foi sendo utilizada como metáfora da expressão musical, alguns artistas começaram a considerar que esta não possuía todos os requisitos que permitissem uma forma de expressão artística visual similar à música. Não será por isso de estranhar que, por exemplo, Baranoff-Rossiné tenha posteriormente substituído o pincel e as tintas por sistemas de projecção de luzes coloridas como o seu *Piano Optophonique* [30], utilizado como instrumento para a construção de “concertos visuais”, claramente influenciados pela linguagem musical e pela própria pintura.

1.6.2 Cinema e Vídeo

A evolução tecnológica ao longo do século XX proporcionou aos artistas que aspiravam a novos meios de expressão artística a concretização das suas ideias através dos órgãos de cor, do cinema, do vídeo e mais tarde do computador. Estes novos meios possuem um denominador comum que os diferencia radicalmente da pintura: a dimensão temporal. Na pintura, ela poderá apenas ser sugerida, mas nunca concretizada. Sendo o tempo um factor essencial para o fenómeno sonoro, torna-se óbvia a importância desta dimensão no resultado final de um produto artístico que se pretenda relacionar com o universo musical.

Com o aparecimento do cinema, muitos pintores começaram a interessar-se pelas possibilidades artísticas inerentes a este novo suporte. A sua exploração culminou com o surgimento do cinema de animação.

Podemos por isso considerá-lo como uma extensão directa da pintura, uma pintura dinâmica que, liberta de quaisquer restrições temporais, é capaz de se desenvolver no tempo. Um destes tipos de animação, a animação abstracta está intimamente relacionada com a música, pois tem sido fortemente influenciada por ela. Isto porque, tendo a música uma natureza abstracta, não representativa, encontra-se em condições privilegiadas de poder inspirar as restantes formas de expressão artística nessa mesma abstracção.

Uma importante referência no âmbito do cinema abstracto é Walter Ruttmann, que no ano de 1921 estreou em Frankfurt a sua curta-metragem *“Lichtspiel Opus I”*, considerado o primeiro filme de animação abstracto a ser apresentado perante uma audiência. Um outro importante animador, Wiking Eggeling, concebeu em 1924 o primeiro filme de animação inteiramente abstracto, intitulado *“Diagonal Symphony”*. Outro nome relevante nesta área é o de Mary Ellen Bute (Figura 6), natural do Texas, EUA. Desde 1934 até 1959, Mary Ellen criou 11 filmes abstractos que foram vistos em inúmeros cinemas dos Estados Unidos, normalmente como curtas metragens introdutórias a outros filmes. Também a sua obra possui uma forte inspiração na música, como é notório até no próprio nome de filmes como *“Rhythm in Light”* (1935), *“Tarantella”* (1941), *“Color Rhapsodie”* (1951), *“Polka Graph”* (1952) ou *“Pastoral”* (1953).

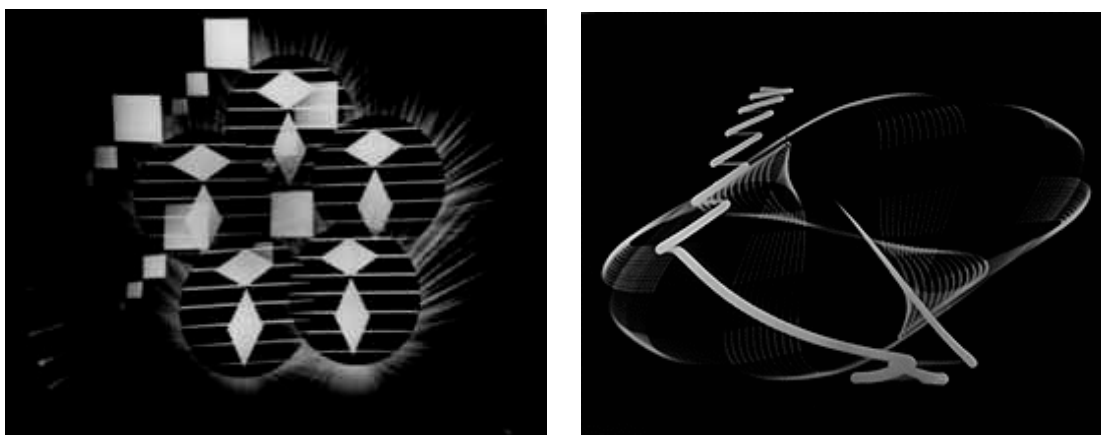


Figura 6 – “Color Rhapsodie” (1951) e “Abstronic” (1954), de Mary Ellen Bute

Apesar da importância dos nomes atrás referidos, Oskar Fischinger (1900-1967) foi provavelmente o cineasta/animador que mais explorou a relação entre a música e a imagem em movimento, através dos seus filmes de animação, construídos ao longo de 30 anos. As técnicas por ele utilizadas vão desde a animação de figuras de barro até à pintura com carvão. Muitas das suas obras denotam uma estrutura rítmica visual com fortes relações musicais. Uma das suas mais importantes colaborações onde este aspecto se encontra claramente evidenciado é o filme de animação *“Fantasia”*, de Walt Disney. Foi ainda um dos pioneiros da técnica da banda sonora óptica. Fischinger pintava formas de onda naquilo a que chamava os *“rolos de som”* (*sound scrolls*) (Figura 7), posteriormente convertidos por mecanismos ópticos para a banda sonora do filme. Na reprodução do filme, Fischinger conseguia assim criar música sintetizada que servia como acompanhamento da sua animação [4]. O animador canadiano Norman McLaren, veio mais tarde a aprofundar esta técnica, conseguindo obter resultados mais elaborados.



Figura 7 – Rolos de som de Oskar Fischinger (esq.) e a sua obra “Allegretto” (1936) (dir.)

Nos anos 40 e 50, os realizadores de cinema John e James Whitney (Figura 8) conceberam filmes abstractos que incluíam efeitos visuais psicadélicos utilizando o comum suporte de 8 ou 35 mm. Alguns dos seus trabalhos podem ser considerados como precursores dos espectáculos de som e luz protagonizados pelas bandas de *rock*, anos mais tarde. No início dos anos 60, com o desenvolvimento tecnológico, conceberam computadores analógicos para animar luzes coloridas, e no decorrer dos anos 70 passaram a utilizar a tecnologia digital. Em 1980, John Whitney publicou uma obra de referência no domínio da Visual Music intitulada *"Digital Harmony - On the Complementarity of Music and Visual Art"* [34]. Contribuiu ainda decisivamente para o aspecto futurista de filmes como *"2001: Uma Odisséia no Espaço"* de Stanley Kubrick. John Whitney esteve sempre a par da evolução tecnológica e contribuiu activamente para a mesma, de tal forma que é considerado por muitos como o *"pai da animação por computador"*. Foi certamente o primeiro técnico/artista a explorar as potencialidades estéticas deste novo meio, e por isso mesmo tornou-se no principal responsável pela legitimação do computador como meio de expressão artística.

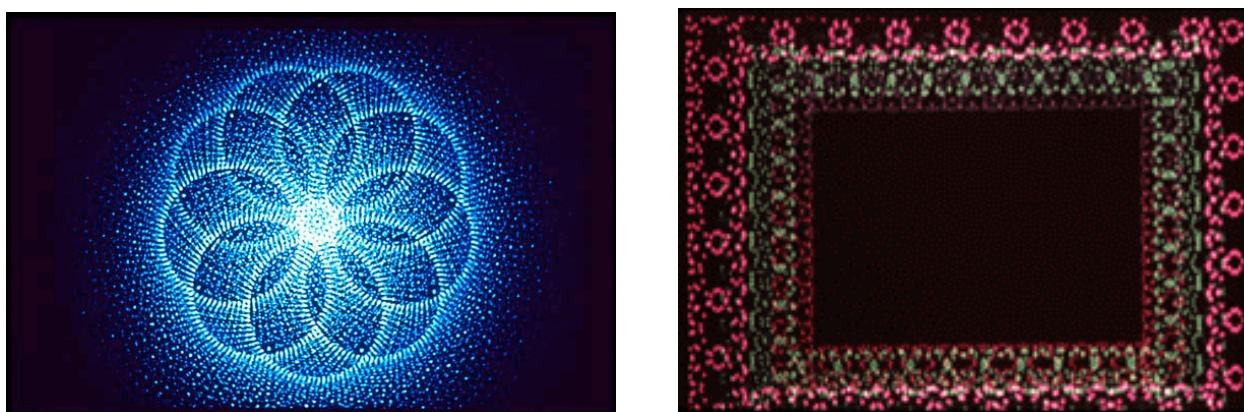


Figura 8 – “Lapis” (1960) de James Whitney (esq.) e “Arabesque” (1975) de John Whitney (dir.)

“As composições na melhor das hipóteses pretendem indicar um caminho em direcção a futuros desenvolvimentos nas artes. Acima de tudo, pretendo demonstrar que a música electrónica e a cor electrónica em movimento se combinam para formar um todo inseparável muito superior que as suas partes.”

Nas duas últimas décadas, com o aumento das possibilidades tecnológicas e com o aparecimento do vídeo no contexto musical, as experiências efectuadas nesta área têm vindo a aumentar. De entre os trabalhos mais recentes, gostaríamos de destacar o vídeo musical do tema “Electronic Performers”, da autoria do duo de música electrónica “Air”. Este filme de animação, totalmente gerado por computador foi realizado por Arnaud Ganzerli, Laurent Bourdoiseau e Jérôme Blanquet em 2003 e tem como temática a representação visual do fenómeno físico da emoção. O filme descreve o percurso de uma emoção, representada pela vibração de ondas sonoras, ao alastrar-se por todo o corpo (Figura 9).

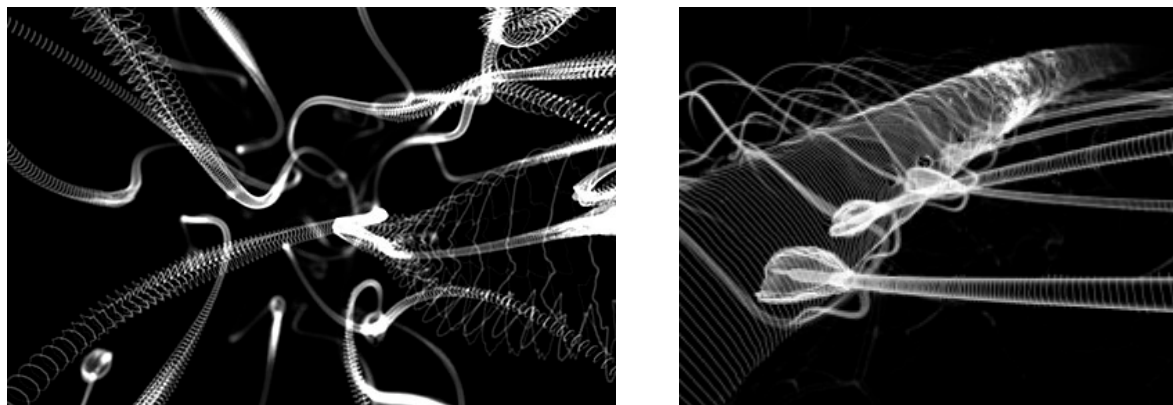


Figura 9 – “*Electronic Performers.*” Música de Air, animação de Arnaud Ganzerli, Laurent Bourdoiseau e Jerome Blanquet

Capítulo II - Oportunidades tecnológicas

2.1 O Áudio Digital

Antes de um evento acústico poder ser digitalizado, este deverá ser convertido num sinal eléctrico analógico. O áudio analógico possui essa designação uma vez que a variação da voltagem ou magnetismo é análoga à variação de pressão acústica do som original. Esta variação, tal como acontece com o som original, é contínua no tempo. A digitalização de um sinal áudio envolve a discretização/segmentação dessa realidade contínua, quer em termos temporais, quer em termos de caracterização da amplitude de cada uma das amostras do sinal [35]. Estamos aqui a partir do princípio que o sinal áudio teve uma origem acústica. No entanto, este pode igualmente ser gerado de forma interna pelo próprio computador, possuindo uma representação digital logo à partida. Embora o processo de digitalização neste caso não exista, a codificação do sinal áudio no domínio digital é em tudo idêntica. Por este motivo, de seguida descrevemos de forma resumida as etapas essenciais no processo de digitalização de um sinal, a amostragem e a quantização, das quais resultam, respectivamente, os dois principais parâmetros do áudio digital: a frequência de amostragem e a resolução das amostras.

2.1.1 Amostragem e Quantização

Através de um mecanismo denominado de *Sample & Hold*, do sinal analógico contínuo são amostrados valores discretos com uma determinada periodicidade. Devido ao facto da própria natureza sonora consistir na variação de pressão acústica no tempo, torna-se necessária a utilização de uma elevada frequência de amostragem (amostras/segundo). No caso do ainda actual *standard* da indústria, o *CD-Audio*, o sinal é definido por 44.100 amostras por segundo, embora formatos mais recentes como o *DVD-Audio*, o *Super Audio CD* ou o *Blu-Ray* possuam a capacidade de trabalhar com frequências de amostragem mais elevadas (48, 96 ou até 192 kHz).

Ao ser utilizada uma frequência de amostragem reduzida, esta pode revelar-se insuficiente para convenientemente representar a onda sonora original, surgindo assim um fenómeno denominado de *aliasing*. Este fenómeno deriva de um problema de inadequação entre a frequência do sinal áudio e a frequência de amostragem utilizada e acontece sempre que a frequência do sinal é superior a metade da frequência de amostragem (*Teorema de Nyquist*) [36]. Por este motivo, conseguimos compreender a razão de, por exemplo, a frequência de amostragem utilizada no *CD-Audio* ser de 44.100 Hz: a frequência máxima perceptível ao ouvido humano – 20 kHz – é inferior a metade da frequência de amostragem utilizada.¹ É de referir que nos conversores AD (analógico-digitais), ainda no domínio analógico, existe um filtro *anti-*

¹ O valor da frequência de amostragem poderia ser qualquer valor acima dos 40.000 Hz. A definição do valor 44.100 Hz está historicamente relacionada com questões de compatibilização com equipamentos de vídeo.

aliasing (LPF¹) que garante que as frequências acima da frequência de Nyquist não passam para a etapa de digitalização [36].

Depois do processo de amostragem, a etapa seguinte consiste na descrição do valor de cada uma das amostras obtidas a partir do sinal analógico segundo o sistema binário. O valor destas amostras é arredondado de forma a ajustar-se às possibilidades de quantização impostas pela resolução utilizada (*bit depth*). No caso do CD-Audio, a resolução utilizada é de 16 bit, resultando em 65536 níveis distintos de amplitude.² No entanto, em termos de produção, o *standard* da indústria é actualmente de 24 bit.

O erro máximo de quantização de um sistema de áudio digital é sempre igual ou menor que metade de 1 LSB (*Least Significant Bit*), logo é independente da resolução utilizada [35]. Sendo este valor fixo, quanto mais níveis de quantização utilizarmos, melhor será a relação entre o sinal digitalizado e o erro de quantização. Dado que o erro de quantização é periódico e está intimamente relacionado com o sinal, provoca distorção harmónica no próprio sinal. É muito mais complicado ignorar auditivamente distorções pontuais do que o ruído constante, uma vez que neste último caso, o nosso sistema auditivo se vai habituando ao mesmo. Uma solução parcial para este problema passa pela utilização de uma técnica denominada de “*Dither*”³. Podemos “aparentemente” eliminar a distorção resultante do erro de quantização adicionando ruído ao sinal. A amplitude do ruído adicionado é sensivelmente equivalente à do próprio erro de quantização. Apesar de não parecer “sensato”, trata-se basicamente de efectuar uma simples troca: distorção harmónica por ruído constante. Esta técnica deve ser utilizada sempre que exista a necessidade de re-quantização, como por exemplo, na conversão de um ficheiro de áudio com uma resolução de 24 para 16 bit (normalmente parte do processo de masterização).

2.1.2 Análise, classificação e descrição de conteúdos áudio

Nesta secção, pretendemos efectuar uma breve referência à vasta área de investigação e desenvolvimento que é a análise, classificação e descrição de conteúdos áudio. Esta tem tido uma grande expansão nos últimos anos devido à crescente utilização de conteúdo audiovisual armazenado em formatos digitais. Para o propósito desta tese, este aspecto revela-se importante, na medida em que, quanto mais informação formos capazes de obter relativamente a um sinal áudio, mais “matéria-prima” teremos para efectuar uma maior correlação entre os domínios sonoro e o visual.

¹ LPF: *Low-Pass Filter*, filtro passa-baixos.

² Segundo a fórmula 2^n , sendo n o número de bits utilizado.

³ Para mais informações acerca da técnica de *Dither*, consultar o excelente artigo de Bob Katz sobre este assunto disponível no seu site, Digital Domain (<http://www.digido.com/>).

Metadata

Devido à crescente utilização do áudio no formato digital, existe actualmente cada vez mais necessidade de catalogá-lo com informação relevante sobre o mesmo. Esta é geralmente anexada ao ficheiro, sendo genericamente referida por *metadata* (*data about data*) [37]. Contrariamente a uma *performance* musical ao vivo, a utilização das funcionalidades proporcionadas pela *metadata* de um ficheiro áudio não se insere num contexto de tempo-real. O conteúdo encontra-se já armazenado e a *metadata* poderá ter sido adicionada de forma manual ou automática, a partir dos mais variados métodos e algoritmos de análise do sinal áudio. Existem inúmeras vantagens proporcionadas pela inclusão de *metadata* num contexto de visualização sonora. Por exemplo, o facto de termos a descrição do género musical incluída nos “ID3 Tags”¹ de um ficheiro *mp3* pode ser útil para permitir ao sintetizador visual a configuração de uma série de parâmetros, tendo em vista uma melhor adequação à música em causa. Existem já alguns *plug-ins* de visualização que exploram determinadas funcionalidades fornecidas pela *metadata* ao incluir o nome da música, álbum e/ou artista como elementos artísticos da própria visualização. Outros, a partir dessa informação utilizam até as imagens dos álbuns em causa com o mesmo propósito, estejam elas armazenadas localmente ou até na *web*.

MPEG-7

O grupo de trabalho da norma MPEG-7² (ISO³/IEC⁴) tem contribuído imenso para a investigação nesta área do conhecimento. Esta norma, também denominada por “*Multimedia Content Description Interface*,” define um meio de descrever os vários tipos de informação multimédia. Ao contrário das normas anteriores (MPEG-1, 2 e 4), esta não se preocupa com o conteúdo (codificação de áudio e/ou vídeo) mas sim com a descrição do mesmo. Esta informação poderá ser associada aos ficheiros de *media* como *metadata* de forma a potenciar uma melhor catalogação e identificação de conteúdos.

O MPEG-7 proporciona estruturas apropriadas para a representação das características sonoras do conteúdo áudio, e um conjunto de funcionalidades básicas. Nesta norma, existem dois tipos de descritores de conteúdo [38]:

- Descritores de baixo nível – existem 17 descritores que podem ser divididos nos seguintes grupos: básicos, espectrais básicos, parâmetros do sinal, tímbricos temporais, tímbricos espectrais, base espectral e silêncio (Figura 10);
- Descritores de alto nível – estes podem ser descritores ou esquemas de descritores. Estes compreendem as seguintes áreas: assinatura digital, timbre dos instrumentos musicais, melodia, reconhecimento geral e indexação, e conteúdo falado.

¹ <http://www.id3.org/>

² MPEG: Motion Picture Experts Group (<http://www.chiariglione.org/mpeg/>)

³ ISO: International Organization for Standardization (<http://www.iso.org>)

⁴ IEC: International Electrotechnical Commission (<http://iec.ch>).

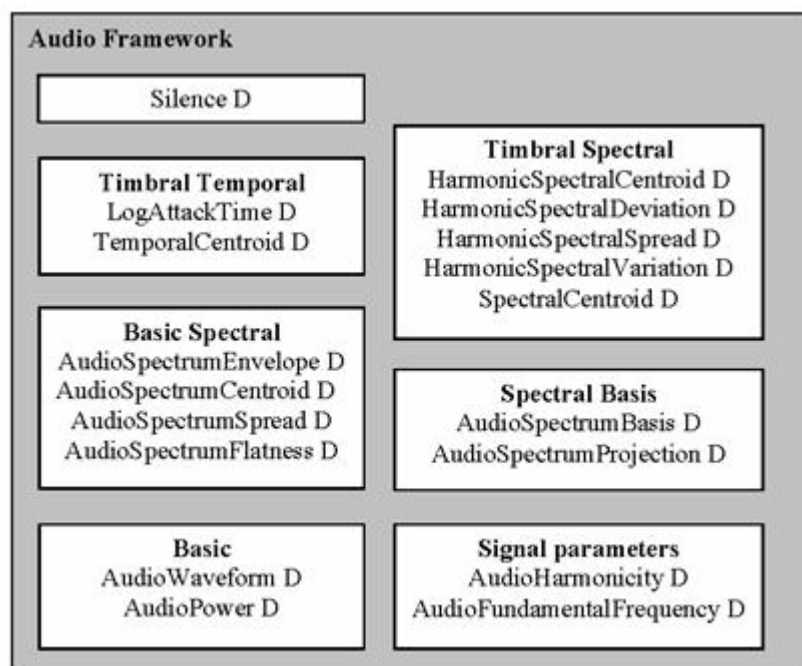


Figura 10 – Descritores áudio de baixo nível da norma MPEG-7 ¹

É importante salientar o facto desta norma não definir o motor de pesquisa que possa vir a utilizar a informação catalogada. De igual modo, não tem o propósito de especificar técnicas que possibilitem a extracção automática de informação a partir da análise do conteúdo multimédia. No entanto, ao centrar a sua atenção na caracterização do conteúdo, fomenta activamente a pesquisa por parte de investigadores e empresas do sector audiovisual nessas mesmas técnicas, uma vez que, desde que possível e fiável, o preenchimento automático dos descritores de conteúdo será preferível a um preenchimento manual.

Investigação em métodos de análise

Sendo esta vasta área de investigação objecto de inúmeras teses de mestrado e doutoramento por si só, torna-se despropositado no contexto desta tese efectuar uma abordagem pormenorizada à mesma. De qualquer forma, julgamos ser importante contextualizar, a título de exemplo, alguns dos esforços que têm vindo a ser feitos neste sentido, uma vez que as vantagens para a visualização sonora são enormes; quanto mais informação musical for possível de extrair de um sinal áudio, maior poderá ser a adequação entre a música e o conteúdo visual gerado.

Dada a importância actual da análise, classificação e descrição de conteúdos áudio, entende-se que o trabalho de investigação sobre este assunto seja muito intenso. Inúmeras empresas, institutos públicos e investigadores no domínio académico têm-se dedicado ao estudo e desenvolvimento de técnicas de análise do áudio de forma a possibilitar a extracção de informação a um nível semântico. Um desses exemplos foi o

¹ Imagem retirada do URL (site MPEG-7): <http://www.chiariglione.org/mpeg/standards/mpeg-7/image061.jpg>

SIMAC (Semantic Interaction with Music Audio Contents)¹, projecto fundado pela Comissão Europeia em Janeiro de 2004 e finalizado em Abril de 2006. A sua função consistiu no desenvolvimento de protótipos para a geração automática de descritores semânticos. Uma referência especial deverá ser feita à *International Society for Music Information Retrieval (ISMIR)*² pelas conferências anuais cujo objecto de estudo incide particularmente nesta mesma área. Os esforços efectuados no desenvolvimento de métodos de análise de áudio enquadram-se nos descritores de alto nível atrás referidos, pois têm o intuito de descrever características como a estrutura e o género musical, a tonalidade, o andamento ou o ritmo. Algumas das características mais recorrentes em termos de análise musical em tempo-real são a identificação dos tempos fortes (*beat tracking*) da música [39], a segmentação temporal e identificação das notas presentes [40], ou a percepção de intensidade da mesma [41]. Tendo em vista uma melhor descrição musical, alguns trabalhos de investigação concentram-se na identificação de segmentos de voz nas músicas [42], de cantores [43] e de instrumentos musicais [44]. Outros na identificação da tonalidade de sinais áudio polifónicos [45], da estrutura musical [46] ou até no carácter da música [47].

Existe imensa actividade de investigação nesta área que cruza inevitavelmente o conhecimento da percepção humana em relação ao som e à música, e o conhecimento científico/técnico que possibilita a programação de sistemas com um complexo poder de análise, com a finalidade de mimetizar a percepção humana.

¹ URL: <http://mtg.upf.edu/static/semanticaudio/>

² URL: <http://www.ismir.net/>

2.2 O MIDI

O MIDI (*Musical Instrument Digital Interface*) consiste num protocolo de comunicação concebido com o intuito de controlar instrumentos musicais electrónicos como sintetizadores, *samplers* e sequenciadores. Esta tecnologia nasceu da necessidade que as várias marcas de equipamentos musicais electrónicos¹ sentiram em construir um *standard* de indústria de forma a compatibilizar os seus produtos. O documento “*MIDI Detailed Specification 1.0*” [48] foi inicialmente redigido em 1982, e em 1983 foi efectuada a primeira demonstração pública na primeira exposição da NAMM (National American Music Merchants). Apesar desta tecnologia ser tão antiga, continua ainda a ser o *standard* da indústria para comunicação entre instrumentos musicais electrónicos. Esta tecnologia define dois aspectos distintos [49]:

1. uma “linguagem”, ou seja, um conjunto de comandos e de parâmetros que possibilitam a comunicação entre equipamentos;
2. as características físicas (cabos, conectores, características eléctricas do sinal) que possibilitam a ligação e respectiva comunicação.

Actualmente, o segundo aspecto tem sido progressivamente substituído por protocolos de comunicação mais modernos como o USB ou o Firewire, sobretudo na interligação de equipamentos MIDI com computadores. Quando a comunicação é feita entre computadores, poder-se-ão utilizar comuns tecnologias de rede. No entanto, o primeiro componente, a definição da linguagem, continua ainda a ser amplamente utilizada. Mesmo numa altura em que a utilização dos chamados “instrumentos virtuais” parece começar a destronar os equipamentos dedicados, estes continuam a ser controlados pelos sequenciadores através da linguagem MIDI.

Um aspecto que julgamos importante reforçar, é o de que o MIDI é uma linguagem de controlo. O conjunto de comandos da linguagem MIDI permite a descrição da informação musical contida na *performance* de um músico e não o som resultante dessa mesma *performance*. Desta forma, um ficheiro ou um *stream* com informação MIDI poderá conter a discretização de praticamente todos os parâmetros de uma determinada *performance* musical.

As principais áreas de acção do MIDI são a composição e *performance* musical, a produção áudio, o controlo e sincronismo entre equipamentos, entre outros. O tipo de equipamentos MIDI existente é extremamente variado, sendo até possível o controlo de sistemas de iluminação num concerto ou o de determinados equipamentos de vídeo.

¹ Com destaque particular para as marcas americanas e japonesas.

2.2.1 Estrutura de uma mensagem MIDI

O protocolo MIDI é um protocolo série, ou seja, envia e recebe um *bit* de cada vez. Estes são agrupados em *bytes* (geralmente entre 1 e 3) que definem comandos ou mensagens. Na prática, o débito máximo de mensagens (informações musicais) que o protocolo MIDI poderá suportar poderá rondar os 1000 a 1500 eventos por segundo. Cada uma destas possui a discretização de determinado evento da *performance* musical. Quando, por exemplo, é premida uma tecla num teclado MIDI, essa acção é descrita numa mensagem constituída por 3 *bytes*. O primeiro *byte* (*command byte* ou *status byte*) é responsável por descrever o tipo de acção; neste caso, uma tecla foi premida (mensagem *Note On*). Os *bytes* seguintes (*data bytes*) caracterizam a acção: qual a tecla premida e a intensidade com a qual foi tocada. A especificação MIDI determina o significado de todos os *status bytes* (excepto os poucos que permanecem indefinidos), bem como o número de *data bytes* que deverão acompanhar cada um dos diferentes *status bytes* e o seu significado [49].

Em cada *byte* MIDI, o MSB (*most significant bit* - bit mais significativo) define o tipo de *byte* – *status byte* ou *data byte*. Isto faz com que a resolução da maior parte das mensagens MIDI seja efectivamente de 7 bit, ou seja, existem 128 possibilidades distintas de quantificação para a discretização dos diversos componentes de uma *performance* musical. Em alguns casos onde esta resolução possa revelar-se insuficiente (ex: *pitch bend*), utilizam-se dois *bytes* combinados para a descrição da mensagem (14 bit).

2.2.2 Canais e Mensagens MIDI

Um aspecto essencial da especificação MIDI consiste na possibilidade de utilização, para determinados tipos de mensagens, de 16 canais distintos de informação. Na prática, esta funcionalidade permitia que através da mesma ligação MIDI, utilizando os conectores MIDI Thru, pudessem ser enviadas mensagens distintas para um máximo de 16 sintetizadores diferentes, cada um programado para receber e reproduzir a informação MIDI de um determinado canal e ignorar a informação dos restantes. Actualmente, a grande maioria dos sintetizadores (e até placas de som para computadores) são já capazes de reproduzir o equivalente a 16 sintetizadores monotímbricos, ou seja, podem tocar toda a informação dos 16 canais de forma autónoma.

As mensagens de canal (*Channel Message*), como o próprio nome indica, são específicas para o canal definido. Assim, podemos controlar, por exemplo, as notas (*Note On/Note Off*), o volume (*Control Change 7*) ou o timbre (*Program Change*) de cada um dos canais MIDI separadamente. Por outro lado, há mensagens cuja utilização pressupõe uma funcionalidade geral, não relacionada com nenhum canal em particular. Exemplos disso poderão ser as mensagens de *MIDI Time Code (System Common)* ou de *Start/Stop (System Real Time)*. Essas mensagens são denominadas Mensagens de Sistema (*System Message*).

Mensagens de Canal

Mensagem de Nota (Note On / Note Off)

Este tipo de mensagem possui 2 parâmetros associados: qual a nota tocada e com que intensidade foi tocada. A mensagem de *Note Off* é em tudo similar mas é de referir que muito poucos equipamentos MIDI possuem a capacidade de medir a velocidade de libertação da nota, logo, este parâmetro é pouco utilizado. A duração de uma determinada nota é definida pela diferença temporal entre o evento *Note On* e o respectivo *Note Off*.

AfterTouch (Mono e Poly)

Estas mensagens destinam-se a definir a pressão aplicada às teclas (ou outros controladores) depois de estas serem premidas. O parâmetro de alteração tímbrica a que este tipo de mensagem está associado é definido pelo sintetizador e, normalmente, varia de som para som. O *AfterTouch* Monofónico possui apenas um valor para todas as notas tocadas, enquanto que o *AfterTouch* Polifónico torna os valores de pressão independentes por nota, possibilitando, por exemplo, definir valores distintos para as várias notas de um acorde.

Controladores (Control Change)

Estas mensagens são extremamente importantes para a versatilidade e expressividade do MIDI, pois permitem a definição de uma série de parâmetros a nível musical (pedais de um piano, expressão, *vibrato*, etc), de mistura (volume, panorâmica, quantidade de efeitos, etc), ou de controlo de outras funções como por exemplo a alteração tímbrica dos sons. Existem 128 controladores distintos (0-127). Cada um destes controladores possui uma resolução de 7 bit, ou seja poderá assumir valores entre 0 e 127. É de realçar que os controladores 0-63 estão emparelhados de forma a ser possível, quando necessário, a implementação de controladores com o dobro da resolução (14 bit), podendo assumir assim valores entre 0 e 16384.

Programas (Program Change)

Uma mensagem de *Program Change*, também denominada de *Patch Change* tem a finalidade de alterar o preset (timbre, patch, instrumento) seleccionado no canal especificado. Depois desta mensagem, todas as notas tocadas passarão a ser ouvidas com o novo timbre escolhido. Um dos “erros” de concepção da especificação MIDI foi o de considerar que os sintetizadores não teriam mais do que 128 timbres distintos, e por isso mesmo este tipo de mensagem possui apenas um data byte. No entanto, para resolver esta limitação, foi reservado um par de controladores denominados de Bank Select (controladores 00/32 - MSB/LSB). Assim, teoricamente, o protocolo MIDI poderá suportar sintetizadores com 2097152 sons distintos (128^3).

Pitch Bend

Esta mensagem permite alterar a afinação de uma nota tocada. É normalmente utilizado um *joystick* ou uma roda com retorno automático ao centro quando libertado/a (afinação correcta). Este tipo de mensagem possui 2 *data bytes* associados, de forma a possuir uma resolução suficiente para que não sejam notórios “saltos” nas alterações de afinação. A quantidade de variação na afinação à qual os valores mínimo e máximo deste controlador correspondem é definido no próprio equipamento, normalmente através do número de meio-tons.

Mensagens de Sistema

System Exclusive (Sistema Exclusivo)

Estas mensagens servem para que instrumentos musicais da mesma marca/modelo possam comunicar entre si, para trocar configurações, ou qualquer outro tipo de dados. Estas mensagens contêm uma implementação MIDI aberta, nem o número de *data bytes* está definido, pois depende da quantidade de informação a ser comunicada. Cada marca implementa estas mensagens como bem entender para o equipamento em questão, visto que estas não necessitam de ser universalmente entendidas. As únicas características definidas na especificação MIDI são as do *status byte* inicial (F0 em hexadecimal) e do *data byte* final (F7).

System Common

Este grupo de mensagens MIDI são denominados de *System Common* por não possuírem informação de canal e por serem comuns a todos os equipamentos numa cadeia MIDI. O tipo de mensagem mais utilizada é provavelmente o MTC (*MIDI Time Code*), implementação do código de tempo SMPTE em MIDI (Horas:Minutos:Segundos:Frames).

System Real Time

Estas mensagens de sistema são utilizadas para efectuar o sincronismo entre equipamentos enquanto estes estão a “tocar”. Elas são constituídas apenas por *status bytes*, que pela sua necessidade de exactidão temporal e por serem constituídos apenas por 1 *byte* podem ser comunicados a qualquer momento, inclusivamente a meio de qualquer outro tipo de mensagem. Como exemplo destas mensagens temos o *MIDI Clock* e os comandos *Start*, *Stop* e *Continue*.

A Tabela 5 que abaixo apresentamos, enuncia os principais tipos de mensagens MIDI existentes.

Tabela 5 – Resumo dos vários tipos de mensagens MIDI

Mensagens MIDI	Descrição
Mensagens de Canal	
Note On / Note Off	Definição de activação/desactivação da nota e respectiva dinâmica
Mono e Poly AfterTouch	Alterações tímbricas mediante a pressão exercida
Control Change	Volume, Panorâmica, Expressão, <i>Portamento</i> , Pedais, ADSR
Program Change	Alteração de programa / configuração
Pitch Bend	Alteração de afinação
Mensagens de Sistema	
System Exclusive	Backup de informação (Bulk Dump)
System Common	MIDI Time Code
System Real Time	MIDI Clock, MIDI Machine Control (Start, Stop, Continue)

2.3 O OpenSound Control

O OSC (*OpenSound Control*)¹ é um protocolo desenvolvido pelo CNMAT (*Center for New Music and Audio Technology*) da Universidade de Berkeley, Califórnia e, à semelhança do protocolo MIDI, tem como propósito possibilitar a comunicação entre computadores, sintetizadores, e outros dispositivos multimédia. A especificação deste protocolo – *OpenSound Control Specification 1.0* [50] – foi redigida em 2002 e, sendo cerca de 20 anos mais recente que o MIDI, revela-se mais apropriado às necessidades actuais, nomeadamente, na utilização de redes de comunicação modernas para a transferência de dados, e na definição de uma linguagem com maior abertura e maior resolução do que a utilizada no protocolo MIDI.

2.3.1 Resumo das características do OSC

A arquitectura do OSC é baseada no conceito de *cliente / servidor*, sendo que qualquer dispositivo que envie pacotes é considerado cliente e qualquer outro que os receba, será considerado servidor. Um dispositivo poderá ser considerado cliente e servidor simultaneamente. O OSC é um protocolo completamente independente da tecnologia física de transporte da informação. Desta forma, torna-se extremamente versátil, podendo-se integrar em praticamente qualquer tecnologia de rede.

No OSC, o método de atribuição de nomes às mensagens é feito de forma dinâmica e aberta, com uma estrutura semelhante à de um URL (ex: “*oscilador/4/freq*”). São suportados vários tipos de argumentos para as mensagens: *strings* ASCII, *ints* e *floats* de 32 bit, 64 bit, entre outros. Este protocolo incorpora um sistema de pergunta (*queries*) para conhecer as capacidades de outro dispositivo OSC de uma forma dinâmica. Possibilita ainda a utilização de funções de *Pattern Matching* (“?”, “*”, “[]”, etc.) com utilização similar à de uma *shell* em UNIX, permitindo assim a especificação de múltiplos destinos para uma única mensagem.

Em termos de gestão temporal, a inclusão de *Time Tags* permite uma alta precisão temporal (abaixo do nano-segundo), utilizando o mesmo formato do protocolo Internet Network Time Protocol (NTP). Contempla ainda a possibilidade do envio de *Bundles* de mensagens (com um *time tag* associado), cuja acção deva ocorrer simultaneamente.

¹ <http://www.opensoundcontrol.org/>

2.3.2 Arquitectura de Rede do OSC

Em termos de arquitectura de rede, os equipamentos ou os diversos componentes de software que utilizam o OSC podem comunicar através de várias formas:

- Multiprocessamento distribuído em *LANs* (cooperação para atingir uma única tarefa)
- *LANs* ou *WANs Peer-to-peer* (comunicação entre dispositivos, mas operação independente)
- Máquina isolada (várias aplicações, *plugins* ou processos residentes na mesma máquina comunicam internamente por OSC).

No contexto da arquitectura protocolar TCP/IP, o OSC encontra-se ao nível da aplicação, utilizando o UDP (User Datagram Protocol) como protocolo ao nível do transporte. Contudo, não existe nenhuma obrigatoriedade em utilizar o UDP; é também possível utilizá-lo com o protocolo TCP. A utilização do protocolo UDP como forma de transporte preferencial no OSC prende-se com o facto de este não estabelecer ligação nem incluir métodos de confirmação da recepção dos datagramas [51]. O UDP é por isso mais utilizado por razões óbvias de rapidez de resposta, visto ser, como o MIDI, um protocolo de comunicação vocacionado para o controlo de equipamentos e de *software* em tempo-real. Neste contexto, a baixa latência proporcionada pelo UDP será mais importante do que a fiabilidade garantida pelo TCP mas, em outros contextos, em que as necessidades sejam distintas, o TCP poderá ser uma escolha acertada.

Uma vez que a própria arquitectura protocolar TCP/IP possui uma grande independência relativamente a infra-estruturas de rede como o tipo de cablagem, os equipamentos ou a topologia, o protocolo OSC liberta-se automaticamente destas preocupações de “baixo nível”, podendo ser aplicada de uma forma simples a praticamente qualquer tipo de rede.

2.3.3 Implementação do protocolo OSC

Alguns exemplos de aplicações ou ambientes de programação com suporte para o protocolo OSC são o Supercollider¹, o Max/MSP², o Pure Data (Pd)³, o Open Sound World (OSW)⁴, o CSound⁵ e o Reaktor⁶. Existem ainda extensões para aplicações como o Adobe Flash,⁷ Director, ou até para linguagens como Perl,

¹ <http://www.audiosynth.com/>

² <http://www.cycling74.com/products/maxmsp>

³ <http://puredata.info/>

⁴ <http://osw.sourceforge.net/>

⁵ <http://www.csounds.com/>

⁶ <http://www.native-instruments.com/#/en/products/producer/reaktor-5/>

⁷ <http://www.benchun.net/flosc>

Python, PHP ou Matlab¹ que proporcionam também a possibilidade de utilização deste protocolo nesses contextos [52].

Em termos comerciais, é ainda clara a proliferação do protocolo MIDI. Enquanto o protocolo OSC não for suportado e adoptado pelas grandes marcas de produtos em *hardware* (Roland, Yamaha, Korg, etc.) bem como pelas marcas que desenvolvem o *software*, receamos que este protocolo, apesar de tecnologicamente superior, não venha a destronar o ainda útil MIDI. E esta implementação, em nosso entender, pode eventualmente depender da criação de uma espécie de modo de emulação do MIDI dentro do OSC, que possa de alguma forma garantir a retro-compatibilidade com todos os equipamentos MIDI produzidos nestes últimos 20 anos. O protocolo MIDI está de tal forma enraizado na indústria que, provavelmente, só uma mudança progressiva, de continuidade e não de ruptura, a nível tecnológico, poderá acabar por o “destronar” em termos comerciais.

Vimos já anteriormente que o MIDI define o “significado musical” dos vários tipos de mensagens como o *Note On/Off*, *Control Change* ou o *Pitch Bend*. Existem apenas duas excepções a esta regra: as mensagens de *System Exclusive* e algumas mensagens de *Control Change* que não estão definidas.² A normalização de parâmetros por parte do protocolo MIDI facilitou a sua adopção por parte da indústria, permitindo uma mais simples comunicação entre equipamentos e *software*.

Sendo o OSC um protocolo muito mais aberto, não define à partida este tipo de dados, deixa a cargo de cada equipamento ou *software* a definição do tipo de mensagens que pode enviar ou interpretar. Este “problema” tem dificultado a adopção do OSC como substituto do MIDI, apesar de tecnologicamente, este protocolo ter todas as potencialidades para o fazer. A definição de um conjunto de parâmetros por parte do protocolo OSC que simule as funcionalidades da especificação MIDI poderia facilitar imenso essa adopção por parte da indústria, sem desvirtuar a possibilidade de uma utilização mais “aberta” e experimental.

¹ <http://andy.schmeder.net/software>

² Estes dois tipos de mensagens não possuem qualquer atribuição de significado por parte da especificação MIDI, podendo ser implementados de forma livre por parte dos fabricantes de *hardware/software*.

2.4 A Síntese Sonora

2.4.1 Breve perspectiva histórica

Desde que Thaddeus Cahill's inventou em 1897 aquele que pode ser considerado o primeiro sintetizador significativo, o “Dynamophone” ou “Telharmonium” [53], muitos avanços foram efectuados até aos dias de hoje. Em 1906, o engenheiro e inventor Lee De Forest patenteou a tecnologia conhecida como o tríodo ou válvula. A sua aplicação inicial esteve confinada às tecnologias de rádio e só em 1915 o seu autor tomou consciência do potencial musical da sua invenção. Assim nasceu o *Audion Piano*, seguindo-se outros instrumentos como o *Theremin* em 1917 e o *Ondes Martenot*, em 1928 [54]. A música electrónica deu aqui os seus primeiros passos. Os pioneiros na utilização destas tecnologias utilizavam componentes tradicionais de áudio como osciladores, filtros, misturadores, entre outros, normalmente utilizados em sistemas de rádio ou em redes telefónicas, mas com uma finalidade de criação musical. Tornava-se assim possível obter sonoridades completamente inovadoras, prontamente aproveitadas por uma série de compositores.



Figura 11 – Tecnologia utilizada para a síntese sonora nos anos 40 (à esquerda) e nos anos 60 (direita)

A utilização do tríodo ou válvula revelou-se fundamental para a síntese sonora até ao aparecimento no início da década de 60 de outra importante invenção neste domínio: o circuito integrado. É nesta altura que surgem nomes incontornáveis da história da síntese sonora, como Robert Moog ou Donald Buchla [53]. A utilização do circuito integrado veio proporcionar o aparecimento dos primeiros sintetizadores (sistemas dedicados para música electrónica) (Figura 11). Estes sintetizadores eram construídos segundo um conceito modular, contendo componentes como osciladores, filtros e misturadores, mas também outros concebidos propositadamente para fins musicais como os geradores de envolvente, moduladores e sequenciadores. A interligação destes componentes era assegurada através de cabos, denominados de *patch cords*, que transmitiam informações através de impulsos eléctricos.

No final dos anos 70, os sintetizadores começaram a utilizar a electrónica digital para o controlo da configuração dos parâmetros sonoros. Os vários módulos passaram de uma forma progressiva a ser incluídos num só equipamento, o sintetizador. Este avanço possibilitou que, entre outros aspectos, a selecção de um determinado timbre (*patch*), anteriormente efectuada através da interligação com cabos dos vários módulos existentes, fosse agora efectuada através do premir de um botão, accionando assim todos os parâmetros registados na respectiva memória [49]. Esta funcionalidade veio proporcionar uma grande vantagem em termos de acessibilidade. A partir dos anos 80, começou a ser utilizada a tecnologia digital não só para o controlo mas também para a própria produção sonora. Talvez o marco mais importante neste tipo de sintetizadores tenha sido o DX7, lançado pela Yamaha em 1983 (Figura 12). Este foi o primeiro sintetizador completamente digital. Outra das características únicas que o evidenciaram foi a utilização de uma tecnologia de síntese inovadora, denominada de FM (*Frequency Modulation*). Esta tecnologia foi desenvolvida por John Chowning, na Universidade de Standford, durante os anos 70. Aliás, o seu artigo publicado em 1973 no *Journal of the Audio Engineering Society*, [55] é considerado por muitos como uma das mais importantes publicações na história da música electrónica. Esta altura coincidiu com o aparecimento de uma tecnologia que está de forma incontornável ligada ao sucesso do sintetizador: o protocolo MIDI, *standard* da indústria, e ainda hoje amplamente utilizado.



Figura 12 – O Yamaha DX7, o primeiro sintetizador completamente digital

Desde o advento da tecnologia digital e do protocolo MIDI, no início dos anos 80, as capacidades e acessibilidade dos sintetizadores têm aumentado consideravelmente. No entanto, contrariamente ao que aconteceu até esta altura, a base tecnológica tem-se mantido estável. Investigadores e marcas de instrumentos musicais electrónicos têm utilizado a tecnologia digital para abordar outros métodos de síntese sonora. Enquanto que até então a grande maioria dos sintetizadores funcionava segundo os clássicos métodos de síntese sonora (subtractiva, aditiva, ou similar), existe agora uma grande variedade de métodos de síntese adicionais como o *waveshaping*, o *wavetable*, a síntese granular e a síntese por modelação física que conferem ao sintetizador uma expressividade sonora ainda mais acentuada.

2.4.2 A síntese sonora e o computador

Até agora focámos a nossa atenção na síntese sonora em *hardware*, ou seja, através de equipamentos dedicados. Relativamente à síntese sonora em *software*, ou seja, através de programas de computador, as primeiras experiências remontam ao ano de 1957, quando o MUSIC I foi desenvolvido por Max Mathews nos Laboratórios Bell e corria num IBM 704. O MUSIC II, lançado um ano mais tarde, permitia já 4 partes polifónicas e o computador mais potente dessa altura necessitava de uma hora de cálculos para efectuar um minuto de música. Este programa foi constantemente melhorado, de acordo com as capacidades dos novos computadores, e culminou no MUSIC V em 1968, a correr num IBM 360. Este projecto foi então aproveitado por outros programadores, entre os quais John Chowning e James Moorer, da Universidade de Standford, onde desenvolveram o MUSIC 10 [54] [56].

Max Mathews lançou um novo sistema em 1970, o GROOVE (Generated Real-time Output Operations on Voltage-controlled Equipment), o primeiro sistema híbrido para síntese musical (controlo digital de um método de síntese analógica). Durante alguns anos existiu esta tendência de utilizar o computador para controlar sintetizadores analógicos, na altura com uma sonoridade mais elaborada. Só no início dos anos 80, com o amadurecimento da tecnologia digital, é que a síntese sonora por computador se começou a desenvolver para lá dos círculos académicos e laboratórios onde estava até então confinada. Em 1982 é lançado o Commodore 64, um computador pessoal acessível com um sintetizador de 3 vozes integrado num *chip*. Este, embora inicialmente idealizado para jogos de computador, foi rapidamente aproveitado também para fins musicais. Por volta de 1984, com o sucesso de sintetizadores digitais como o Yamaha DX7, do protocolo MIDI e do aparecimento de computadores como o Apple Macintosh, o futuro da música por computador estava de certa forma garantido.

Desde esta altura, a utilização da síntese sonora em *software* tornou-se cada vez mais utilizada através de programas como o Csound,¹ entre outros. Este programa é ainda hoje uma referência em termos de investigação em áreas relacionadas com a síntese sonora, mas existe actualmente um conjunto de programas igualmente capazes e mais acessíveis em termos de *interface* a compositores e músicos. Para além dos simples instrumentos virtuais² que simulam a arquitectura de uma determinada máquina, existem ainda ambientes visuais de programação completamente configuráveis como o Reaktor e o Max/MSP.

É conveniente referir que nos últimos anos, devido ao aumento exponencial das capacidades de processamento dos computadores pessoais, tem-se assistido a uma tendência em favorecer a utilização de sintetizadores e *samplers* implementados em *software* em detrimento do *hardware* dedicado utilizado no

¹ <http://www.csounds.com/>

² Termo utilizado para os sintetizadores, *samplers*, ou outros instrumentos baseados em *software*.

passado. Esta nova realidade começou por se fazer sentir em contextos de estúdio, mas a utilização dos instrumentos virtuais em actuações ao vivo é hoje prática cada vez mais comum. Para esta tese de mestrado, este facto revela-se extremamente importante, dado que a abordagem à síntese sonora através de *software* possibilita à partida um maior grau de abertura em termos de programação sonora e, em alguns casos, de maiores possibilidades de interacção com outras aplicações do domínio audiovisual.

2.4.3 Componentes de um sintetizador

A base de qualquer sintetizador, independentemente do método de síntese utilizado, é o elemento gerador de sinais, sejam estes sons simples ou complexos, gerados sinteticamente ou através de amostras de áudio digital. Associado a este elemento, normalmente designado por “fonte”, poderemos ter “modificadores”, dispositivos responsáveis pela alteração do sinal áudio de alguma forma. Alguns dos modificadores mais frequentes são os filtros, os geradores de envolvente e os osciladores de baixa frequência [56]. Neste espaço não pretendemos descrever estes elementos de forma exaustiva, apenas pretendemos referir as suas funcionalidades básicas, uma vez que estes elementos serão utilizados na construção do modelo que nos propomos realizar.

Fonte

A fonte, também denominado de gerador ou até oscilador, é o elemento responsável pela matéria-prima sonora de todo o sintetizador. Os sons gerados podem ser baseados em formas de onda simples como a sinusoidal, triangular dente de serra e quadrada, ou podem ser baseados em amostras de sons com um espectro sonoro mais complexo. O método de síntese sonora de determinado sintetizador é definido fundamentalmente por este componente. Por exemplo, no caso da síntese sonora subtractiva, utilizar-se-ão formas de onda harmonicamente mais complexas, enquanto que num método de síntese aditiva, poderão ser utilizadas múltiplas ondas sinusoidais.

Filtros

Os filtros afectam o sinal de entrada, através da atenuação de determinada zona de frequências, mas também através da amplificação. Os filtros mais conhecidos são:

- filtro passa-baixo (LPF): filtra todo o conteúdo acima da frequência de corte
- filtro passa-alto (HPF): filtra todo o conteúdo abaixo da frequência de corte
- filtro passa-banda (BPF): junção dos dois filtros anteriores. Apenas deixa passar uma zona de frequências restrita.
- filtro *notch*: filtra de forma inversa ao anterior; (é antes rejeita-banda).

Os principais parâmetros associados a um filtro são a sua frequência de corte e a ressonância. A primeira consiste na frequência a partir da qual o sinal passa a decair até ser completamente filtrado. O decaimento é

normalmente referenciado através do valor dB/oitava. Um valor alto como “36” traduz-se num decréscimo acentuado, enquanto que um valor baixo como “6” significa que a actuação do filtro será mais progressiva. O segundo parâmetro, a ressonância ou “factor Q” consiste na amplificação do sinal na zona da frequência de corte, acabando por funcionar como um ressoador [53]. Uma importante funcionalidade dos filtros no contexto da síntese sonora é o chamado *pitch tracking* ou *keyboard scaling*. Esta funcionalidade permite que a definição da frequência de corte do filtro seja alterada em função da nota que está a ser tocada pelo sintetizador, permitindo uma homogeneidade na estrutura harmónica ao longo da tessitura do instrumento.

Gerador de Envolvente

O comportamento de qualquer som ao longo do tempo sofre habitualmente alterações. Estas poderão ser reflectidas na dinâmica, na afinação ou até no próprio timbre. A forma como estas características variam ao longo do tempo pode ser definida por aquilo a que chamamos a envolvente ou perfil sonoro. Ao analisarmos o som produzido pela maioria dos instrumentos acústicos podemos concluir que existem quatro momentos extremamente importantes: o ataque, o decaimento, a sustentação e a extinção [11]. Cada uma destes parâmetros corresponde ao que se denomina por transitórios da envolvente sonora ou transitórios ADSR¹ (ver Figura 13). No contexto da síntese sonora, não estamos limitados à definição destes quatro momentos; existem aplicações que permitem a definição de pontos temporais adicionais, ou até a definição desta envolvente através de curvas relativamente complexas, mas na maior parte dos casos, a definição destes 4 parâmetros (ADSR) revela-se suficiente.

- Ataque:** este é o momento inicial no qual o instrumentista acciona o instrumento. No ataque estão contidas as principais informações que nos fazem reconhecer um instrumento musical; se retirarmos a porção do ataque ao som de um instrumento acústico, poucos músicos o conseguirão identificar.
- Decaimento:** este momento segue-se ao ataque, e é considerado um período de transição entre este e a manutenção. Normalmente, dadas as características mais agressivas do ataque na maioria dos instrumentos, neste momento costuma haver uma diminuição na amplitude sonora.
- Sustentação:** a sustentação do som, também chamada de estado estacionário, possui habitualmente o período de tempo mais longo destes quatro. Nesta fase podem ocorrer apenas ligeiras alterações sonoras como ondulações e oscilações, mas o som é essencialmente estável. Este momento pode-se prolongar indefinidamente enquanto o músico o deseje, como no caso de um órgão, ou pode ter uma duração finita, como no caso de um piano ou guitarra.
- Extinção:** Este momento coincide com a “libertação” das notas no instrumento. Nesta altura, o som começa a decrescer de uma forma acentuada até deixar de ser ouvido. Este parâmetro define o tempo ocorrido desde a libertação da nota até à extinção completa do som.

¹ ADSR: “Attack”, “Decay”, “Sustain”, “Release”

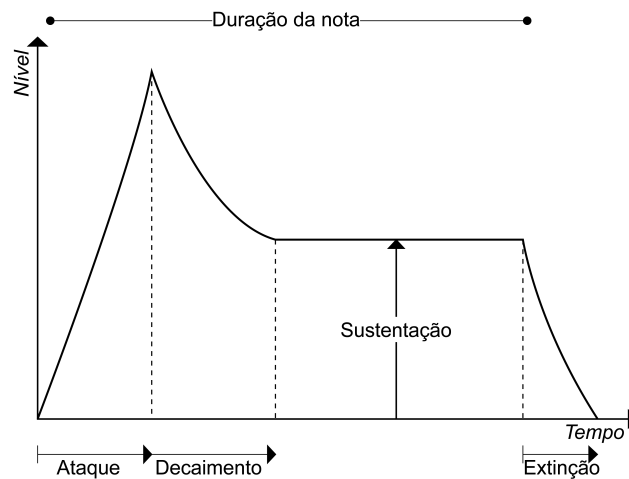


Figura 13 – Comportamento de uma envolvente ADSR

Oscilador de baixa frequência (LFO)

Os osciladores de baixa frequência, normalmente referenciados pela sigla “LFO” (*Low Frequency Oscillator*), são osciladores com formas de onda relativamente simples, similares às utilizadas tradicionalmente na síntese sonora subtrativa e mais algumas associadas especificamente a este contexto (*sample & hold*, ruído, aleatória, etc.). Contrariamente aos osciladores geradores de frequências áudio, aqui a geometria da forma de onda é mais importante que o seu conteúdo harmónico. Outra diferença fundamental aqui presente é a de que estes osciladores vibram a frequências muito baixas, abaixo do limiar de audibilidade (desde os 0 Hz até ao 10 Hz, normalmente). Não sendo estes osciladores audíveis, contribuem para o som final de um sintetizador através da modulação de algum parâmetro como a variação cíclica do volume (proporcionando um efeito de tremolo), da afinação (vibrato), a abertura de um filtro ou a localização do som em termos de panorâmica. Qualquer parâmetro pode em teoria ser modulado através deste mecanismo de variação cíclica.

Processamento de Efeitos

Outros modificadores cada vez mais comuns nos sintetizadores são os processadores de efeitos. Estes podem conter efeitos de *chorus*, *reverb*, *flanger*, *phaser*, *delay*, ou distorção, entre outros [53]. Podem ainda incluir o chamado processamento de dinâmica, como compressores / limitadores, etc. As unidades de processamento de efeitos começaram por ser adicionadas externamente aos sintetizadores na década de 70, de forma a melhorar a sonoridade dos mesmos. Na década de 80 o processamento de efeitos começou a ser incluído nos sintetizadores e nos anos 90, com o amadurecimento da tecnologia digital, passou a ser considerado como parte integrante da estrutura sonora dos sintetizadores, podendo por isso ser hoje entendidos também como “modificadores”.

2.4.4 Métodos de Síntese Sonora

Síntese aditiva

O seu princípio de funcionamento encontra-se intimamente relacionado com o Teorema de Fourier, que em termos práticos diz que qualquer onda complexa pode ser “sintetizada” a partir da junção de um número infinito de ondas sinusoidais de diferentes frequências, amplitudes e relações temporais (fase) [53]. O instrumento baseado em síntese aditiva mais conhecido será provavelmente o órgão. Neste tipo de síntese, a construção do timbre é efectuada através da definição do nível de cada componente harmónica (f_1 , f_2 , f_3 , etc)¹. Em teoria esta forma de síntese poderia ser a única necessária, atentando ao teorema acima mencionado; existem no entanto razões óbvias para considerar outros métodos, por uma questão de acessibilidade (incomportável a configuração de todos os harmónicos de sons muito complexos), capacidade e preço, visto que um sintetizador necessita de possuir um oscilador por cada parcial ou harmónico gerado.²

Síntese subtractiva

Esta é a forma de síntese sonora provavelmente mais utilizada. As ondas emitidas pelo gerador contêm uma determinada estrutura espectral (normalmente ondas com uma representação gráfica simples como a sinusoidal, triangular ou dente de serra), que será “esculpida” através da utilização de filtros. Estes possibilitam a atenuação, filtragem completa ou até em determinados casos, a amplificação de determinadas zonas de frequência. O modificador “filtro” possui neste método de síntese um papel fundamental. A síntese subtractiva é provavelmente a mais utilizada em sintetizadores analógicos e os seus conceitos são ainda hoje aplicados a outros métodos de síntese sonora.

Síntese por modulação de anel (RM), frequência (FM) e amplitude (AM)

A síntese por modulação baseia-se sempre em dois componentes essenciais: uma onda denominada por “portadora” (*carrier*) e outra, de “moduladora” (*modulator*), utilizada para alterar alguma característica da onda portadora. Ambas as ondas encontram-se normalmente em zonas de frequência audível, facto que faz com que se possam alterar consideravelmente as características tímbricas do som. Esta abordagem, para além de ser de simples implementação tecnológica, pois necessita de poucos recursos, traz imensas possibilidades de variação tímbrica, comprovado pela popularidade que os Yamaha DX7 obtiveram (síntese FM) [55].

¹ As diferenças de fase serão à partida ignoradas, uma vez que o nosso sistema auditivo é relativamente insensível a tais mudanças e não as interpreta tímbricamente de uma forma muito acentuada.

² Isto para o caso da forma de onda ser a tradicional sinusoidal, a mais utilizada neste tipo de síntese.

Síntese por *Wavecycle* e *Wavetable*

Estes métodos baseiam-se na utilização de tecnologia digital para proporcionar à fonte uma maior diversidade tímbrica. Em vez das tradicionais formas de onda utilizadas em sintetizadores analógicos, a tecnologia digital torna possível a utilização de formas de onda complexas, geradas sinteticamente ou gravadas e armazenadas numa memória. Estas podem ter a duração de um único ciclo (*single-cycle*), assemelhando-se a um típico oscilador, mas com mais possibilidades de forma de onda, ou vários ciclos (*multi-cycle*). O método *wavetable* utiliza vários *wavecycles*, comutando entre estes de uma forma controlada, alterando também o timbre sonoro.

Síntese por *Waveshaping*

Este método consiste na distorção de uma forma de onda através da utilização de um amplificador não-linear. Estes possibilitam o controlo da forma como o amplificador processa os sinais de entrada. O utilizador altera a sonoridade do sintetizador através de um gráfico da função de transferência, tornando possível alterar o contorno da forma de onda utilizada como fonte do sintetizador. Visto que esta técnica gera frequências adicionais, é normalmente alimentada por um oscilador com formas de onda simples, como a sinusoidal.

Síntese Granular

A síntese granular tem bases radicalmente diferentes de todas as outras. Em vez do modelo fonte/modificador, este método de síntese digital utiliza o conceito de grãos, pequenas amostras sonoras, normalmente com a duração de 20 ou 30 milissegundos, que são reorganizados no tempo e manipulados através de parâmetros como o número de grãos existentes num determinado período de tempo, a componente espectral e a amplitude dos mesmos. Para que a reprodução destes grãos não gere ruídos, a cada grão é aplicado um envelope de amplitude (*fade-in, fade-out*). Dada os seus requisitos de processamento e o seu carácter mais exploratório, este método de síntese não tem sido utilizado em instrumentos musicais comerciais, tendo até agora estado mais confinado ao universo da síntese por *software*.

Síntese por Modelação Física

A construção da estrutura do sintetizador baseia-se aqui na construção de modelos matemáticos que simulem o comportamento de cada um dos componentes dos instrumentos acústicos. Assim, o som gerado será definido pela forma de excitação do instrumento (percutido, dedilhado), pelo tipo de instrumento (corda, tubo, pele), pelas propriedades ressoadoras do instrumento (tamanho, geometria, material), forma de irradiação sonora, entre outras características. Através desta abordagem matemática, torna-se até possível construir instrumentos virtuais impossíveis de simular em termos físicos. Este tipo de síntese, devido à grande necessidade de recursos de processamento que exige, possui actualmente as condições ideais para se poder desenvolver.

2.5 A Imagem Digital

Nesta secção, pretendemos, de forma contida, enunciar apenas as principais características de uma imagem no formato digital e as suas especificidades no que concerne à tese em causa. São aqui abordadas a utilização de gráficos *bitmap*, gráficos tridimensionais através da API OpenGL, e algumas questões essenciais para a compreensão da animação e vídeo digitais.

2.5.1 Gráficos *Bitmap*

Dada a necessidade de discretização de qualquer sistema digital, esta é provavelmente a tecnologia de imagem mais utilizada e consiste, à semelhança do que acontece no áudio digital, na segmentação de uma realidade contínua (Figura 14). Neste caso, as imagens são definidas através de uma matriz de pontos, denominados de *pixel* (abreviatura de *Picture Element*) [57]. A resolução da imagem será dada pelo número de *pixels* na horizontal e pelo número de *pixels* na vertical. Por exemplo, a placa gráfica em grande parte dos computadores equipados com monitores de 20 ou 22 polegadas estará configurada com uma resolução de 1680 x 1050.

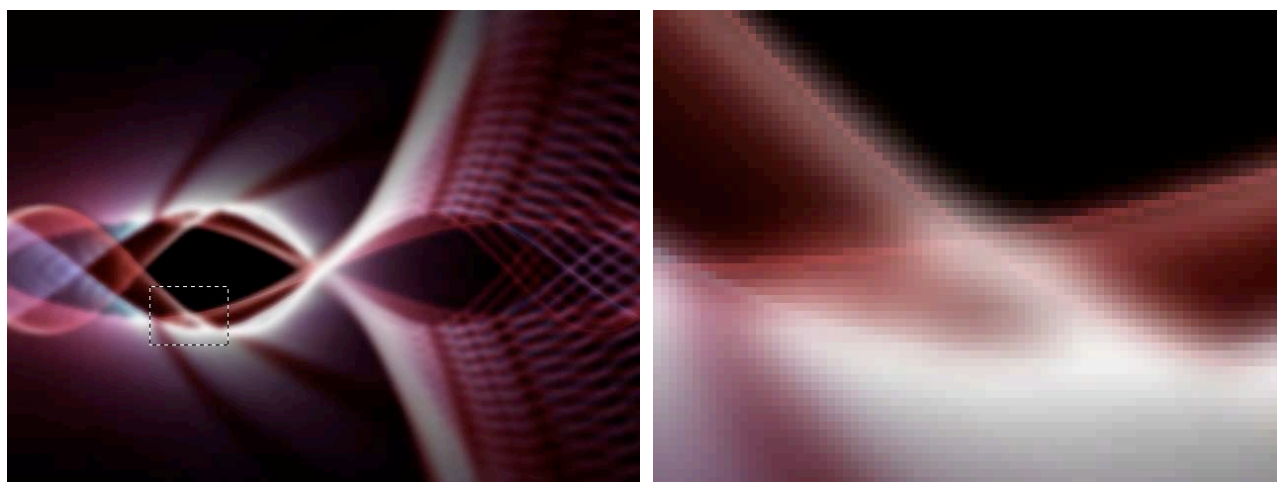


Figura 14 – À esquerda, uma imagem *bitmap* com uma área seleccionada, visível à direita com um zoom de 800%

Cada um destes *pixels* será caracterizado com uma determinada resolução ao nível da cor. A esta resolução é habitual dar o nome de *profundidade de cor* (*bit depth*). Um sistema que utilize 1 *bit* para caracterizar cada *pixel* terá, logicamente, apenas duas cores passíveis de ser representadas (ex: 0 – Preto, 1 – Branco). Quanto mais *bits* forem utilizados para quantificar a cor de cada *pixel*, mais possibilidades teremos. Actualmente, a definição mais comum é a de 24 ou 32 bit, divididos pelas cores primárias da teoria aditiva, conforme apresentado na Tabela 6.

Tabela 6 – Distribuição de *bits* pelos canais de cor

Nº bits	Red	Green	Blue	Alpha
24	8	8	8	
32	8	8	8	8

Cada canal possui uma definição de 8 *bit*, logo, teremos uma gradação de 256 níveis distintos para cada. O canal *alpha* é responsável pela definição do nível de transparência de cada *pixel*, [58] usado particularmente em contextos de composição de imagem (ou de vídeo).

Visto que no contexto desta tese, a utilização da imagem digital será sempre com o intuito de a mostrar em monitores CRT, LCD ou projectores, o modo de cor utilizado será o RGB¹ e não o CMY(K)², utilizado para impressão, onde lidamos com pigmentos e não directamente com o fenómeno luminoso. As unidades DPI (*dot per inch*) ou PPI (*pixel per inch*), são aqui também completamente irrelevantes, pois o tamanho efectivo de visualização não constitui um parâmetro controlável. Interessa-nos apenas a resolução efectiva de apresentação e não o tamanho que a imagem poderá vir a “ocupar” em termos físicos; a visualização do nosso trabalho tanto pode ser feita por um convencional monitor de 15 polegadas ou por um projector de vídeo numa tela de grandes dimensões. Sempre que lidamos com vídeo digital ou aplicações multimédia estas unidades não possuem, portanto, qualquer relevância.

2.5.2 OpenGL

O OpenGL (*Open Graphics Library*) é uma API (*Application Programming Interface*) que define um conjunto de recursos e funcionalidades apropriados para o desenvolvimento de aplicações interactivas em 2D ou em 3D. Foi concebido em 1992 pela *Silicon Graphics Inc.* e tornou-se a API de referência nesta área. Nesta altura, foi constituída um consórcio independente denominado de OpenGL ARB (*Architecture Review Board*) formado por várias marcas representativas da indústria. Este consórcio é responsável pelo desenvolvimento e/ou aprovação de melhoramentos à especificação, mantendo esta API o mais actual possível de forma a dar resposta às cada vez mais exigentes necessidades do mercado [59].

O OpenGL acelera imenso o desenvolvimento de aplicações ao incorporar um vasto conjunto de funcionalidades ao nível da geometria (linhas, polígonos, NURBS, etc), iluminação, *rendering*, *shaders*, mapeamento de texturas, cor e transparência, *anti-aliasing*, *blending*, etc. Estas funcionalidades poderão ser

¹ Red, Green, Blue; cores primárias da teoria aditiva da cor, relacionada com o fenómeno luminoso (utilizada para visualização em televisores e monitores LCD, projectores, etc.)

² Cyan, Magenta, Yellow; cores primárias da teoria subtractiva da cor, relacionada com tintas e pigmentos (utilizada para impressão). O K é um canal separado com tinta preta (*black*), de modo a poder imprimir o preto sem ter de recorrer à junção das 3 cores primárias (CMY).

extremamente úteis a áreas tão vastas como a animação 3D, jogos de computador, realidade virtual, *broadcast*, CAD/CAM/CAE, visualização médica, ou apresentações multimédia.

Esta tecnologia possui uma óptima qualidade e desempenho, uma vez que tem suporte directo ao nível do GPU (*Graphics Processing Unit*) da vasta maioria das placas gráficas (Nvidia, ATI, Matrox, etc). Este facto constitui uma enorme vantagem, uma vez que, como é sabido, actualmente os GPUs das placas gráficas são superiores em capacidade de processamento gráfico do que os próprios CPUs dos computadores (embora não tão versáteis).

O OpenGL é actualmente o *standard* da indústria nesta área, sendo suportado em quase todas as plataformas de *hardware* e sistemas operativos existentes. Pode ser utilizado em linguagens de programação como o C, o C++ ou o Java, entre outras. Ao nível do desenvolvimento de *software*, existem várias vantagens na utilização desta API:

- *Standard* de Indústria - a única API gráfica aberta e multi-plataforma
- Estável - as adições à especificação são bem controladas, garantindo sempre a retro-compatibilidade.
- Fiável e portátil - os resultados obtidos são os esperados, independentemente do sistema operativo ou *hardware* utilizado.
- Evolutiva - permite novas inovações através do recurso a extensões.
- Escalável - pode ser utilizada em equipamentos de consumo até às *workstations* mais potentes.
- Fácil de usar - permite a abstracção do *hardware* na fase de programação.
- Bem documentada - existe imensa bibliografia e código fonte disponível, tornando esta API extremamente acessível a todos os níveis.

Extensões

Apesar do OpenGL ARB (*Architecture Review Board*) definir uma especificação completa da API, qualquer marca possui total liberdade para lhe acrescentar novas funcionalidades através do conceito de *extensões*. Por exemplo, através de extensões, determinadas tarefas podem ser executadas em *hardware* dedicado, ou no CPU do computador, ou em ambos, tornando esta tecnologia extremamente flexível. Ao utilizar este mecanismo de extensões, as marcas de placas gráficas podem diferenciar os seus produtos da concorrência, ao desenvolver determinadas extensões que permitam novas funcionalidades com aceleração gráfica directamente no GPU.

A API do OpenGL está em evolução permanente. Revisões à especificação ocorrem regularmente, e o mecanismo de extensões tem aqui um papel muito importante. À medida que determinada extensão vai ganhando aceitação e relevância, é considerada a sua inclusão na especificação formal do OpenGL. Este processo permite que o OpenGL se desenvolva de uma forma controlada e, simultaneamente, inovadora.

Shaders

A versão 2.0 do OpenGL¹ tem como principal novidade a inclusão da *OpenGL Shading Language* (GLSL)² e das respectivas APIs para criar, gerir e utilizar esses mesmos *shaders* programáveis. Esta linguagem baseada em ANSI C apesar de já existir desde a especificação 1.4 do OpenGL, não fazia parte integrante da mesma, funcionando como uma extensão.

A tendência da Indústria em termos de *hardware* gráfico tem sido a de substituir funcionalidades fixas (Figura 15) pela possibilidade de programação, nomeadamente nas áreas de processamento de vértices e de fragmentos (*vertex*, *fragment*)³. Esta linguagem foi concebida com o intuito de permitir que os programadores possam definir o processamento que ocorre nestas etapas da *pipeline* do OpenGL. É a este código, unidades independentemente compiláveis escritas nesta linguagem que damos o nome de *shaders*. Um *programa* é um conjunto de *shaders* compilados e interligados [60].

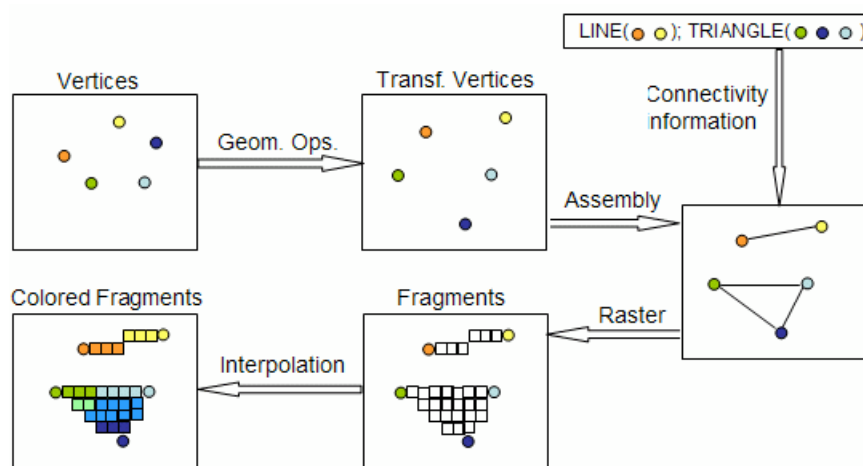


Figura 15 – Esquema da funcionalidade fixa da *pipeline* do OpenGL⁴

Qualquer placa gráfica actual da NVidia, da ATI ou mesmo da Intel possui esta versatilidade, permitindo correr no GPU processos que anteriormente teriam de ser executados pelo CPU principal. A vantagem em termos de *performance* torna-se muito clara. De facto, o aumento das capacidades de processamento paralelo das placas gráficas consegue suplantar claramente as do processador principal no que diz respeito a quase todo o tipo de processamento gráfico, visto que este, com a sua utilização muito mais abrangente, não se encontra optimizado para a especificidade aqui requerida. No contexto desta tese, esta característica vem

¹ <http://www.opengl.org/documentation/specs/version2.0/glslspec20.pdf>

² <http://oss.sgi.com/projects/ogl-sample/registry/ARB/GLSLangSpec.Full.1.10.59.pdf>

³ O processamento de vértices refere-se à alteração da geometria dos modelos OpenGL, enquanto que o processamento de fragmentos está relacionado com o processamento da cor de cada *pixel*.

⁴ Imagem retirada do URL: <http://www.lighthouse3d.com/opengl/glsl/images/visualpipeline.gif>

permitir libertar o processamento do CPU para executar determinadas tarefas que não seja possível implementar no GPU ou cujo tipo de cálculo não se adequa tanto a este.

2.5.3 Vídeo e Animação

Como acontece com qualquer tipo de informação existente no domínio digital, o movimento tem necessariamente de ser reduzido a uma sucessão de eventos discretos, neste caso, imagens estáticas, denominadas de quadros (*frames*). A frequência com que esta discretização de uma realidade que na sua essência é contínua, ininterrupta, possui também, à semelhança do que acontece no áudio digital, as suas limitações. No início do século XIX, o físico Peter Mark Roget chegou à conclusão que uma sequência de imagens estáticas sugere a sensação de movimento quando apresentada com uma frequência acima das 12 imagens por segundo [61]. Logicamente, esta ilusão de movimento será progressivamente mais eficaz à medida que o número de imagens por segundo aumenta. Contudo, existem questões práticas no âmbito da implementação tecnológica que restringem a utilização de frequências de amostragem muito elevadas nas áreas do vídeo e animação, pois esta utilização irá provocar necessariamente uma sobrecarga, tanto em termos de processamento como em termos de armazenamento da informação visual.

Analisando os principais *standards* de vídeo existentes na indústria audiovisual (Tabela 7), podemos confirmar que não serão necessárias muitas imagens por segundo (*framerate*) para conseguir ludibriar o nosso mecanismo visual a interpretar essa sucessão de imagens estáticas como movimento.

Tabela 7 – Número de imagens por segundo (*framerate*) utilizados pelos principais sistemas de imagem [62]

	PAL	NTSC	Cinema
Framerate	25	29,97	24

A afirmação atrás feita pode ser considerada uma meia verdade. Com estes valores, a sensação de movimento é garantida mas não necessariamente de uma forma fluida. Analisando o modo como as imagens são captadas no cinema ou através de uma câmara de vídeo, podemos verificar que cada imagem é constituída pelo somatório das “imagens” contidas num certo intervalo de tempo. Desta forma, não existe nenhum momento temporal que não esteja representado num determinado quadro. Na prática, o resultado visual deste procedimento resulta no fenómeno denominado por *motion blur*, ou seja, a desfocagem da imagem motivada pelo movimento da câmara e/ou dos objectos (ver Figura 16). Esta desfocagem auxilia a nossa percepção quanto à fluidez do movimento pois assemelha-se à forma como a própria visão humana capta as imagens.



Figura 16 – Efeito de desfocagem motivada pelo movimento (*motion blur*)

Existe porém uma diferença fundamental a ter em conta quando lidamos com imagem sintética, gerada pelo computador. Seja num jogo de computador baseado na geração de gráficos tridimensionais em tempo-real ou num sistema de *render 3D* a calcular as imagens relativas a um filme de animação, as imagens reflectem apenas um determinado momento temporal, não um intervalo de tempo. Desta forma, estas estarão sempre nítidas, resultando numa menor fluidez de movimento. Se no caso do *render* de cenas tridimensionais em aplicações como o Maya, sem requisitos de tempo-real, é possível processar a imagem com elaboradas técnicas de *motion blur*, no caso do *render* em tempo-real, a solução mais razoável será a de gerar imagens com a maior frequência possível para que mais momentos temporais possam ser representados visualmente.

Esta problemática ganha particular relevância na área de abordagem desta tese, onde a natureza sintética da imagem pode provocar essas descontinuidades temporais. Dadas as suas características de *performance* em tempo-real, é necessário que a rapidez de resposta e a fluidez de movimentos da visualização sonora sejam garantidas. Esta é a razão pela qual, à semelhança de um jogo de computador e contrariamente ao cinema e vídeo, o número de imagens por segundo não é à partida estipulado. Este valor dependerá directamente da capacidade de processamento do CPU e do GPU do computador, sendo que, quanto maior for o *framerate*, melhor.

Uma das características incontornáveis da animação e do vídeo digital é, sem dúvida, dada a grande quantidade de informação gerada, a da codificação/compressão de informação [58]. No entanto, no contexto desta tese, este aspecto não possui muita relevância, uma vez que as imagens são geradas e visualizadas em tempo-real. Assim, não existe qualquer necessidade de compressão de informação, seja por motivos relacionados com a largura de banda disponível, como acontece num contexto de *broadcast* ou de difusão pela *web*, seja pela ocupação de espaço de armazenamento em suportes como discos duros, CDs ou DVDs.

2.6 Exemplos de Visualização Sonora

Neste espaço pretendemos abordar alguns exemplos de visualização sonora aplicados a uma área de utilização no âmbito do comum consumidor: os visualizadores de música, integrados na forma de *plug-in* nas *jukeboxes* como o iTunes, o Winamp ou o Windows Media Player. Esta tem sido uma das áreas onde a visualização criativa do som mais se tem expandido. O aspecto de muitos dos visualizadores existentes possui uma enorme qualidade e grande variedade de resultados. Pela observação por nós efectuada no decorrer deste trabalho, estas virtudes foram de alguma forma ocultadas por um aspecto que, dado o carácter da nossa tese, nos parece de enorme relevância: a escassa relação entre a acção musical e a reacção visual. Na maior parte dos casos nota-se a existência de dois discursos distintos: o sonoro e o visual. Apenas alguns elementos como a representação visual da forma de onda ou do espectro dos sinais áudio e a detecção básica dos tempos fortes da música permitem uma ligeira aproximação dos dois. A pouca quantidade de características sonoras com uma representação visual leva a que não sejamos capazes de experienciar um único objecto audiovisual mas sim a soma de duas experiências distintas. Uma vez que as crescentes possibilidades de análise de sinais áudio permitem actualmente o reconhecimento de um vasto conjunto de características sonoras, seria desejável que estas pudessem ser aplicadas também nesta correlação audiovisual.

Podemos afirmar que uma das aplicações pioneiras neste domínio foi, no início dos anos 90, o *Cthugha*,¹ escrita por Kevin "Zaph" Burfitt para o sistema operativo MS-DOS. Esta aplicação, anos antes da proliferação das *jukeboxes* digitais e do formato *mp3*, permitia a visualização criativa de áudio, apesar de estar limitada aos recursos técnicos da altura (placas gráficas com baixa resolução e um máximo de 256 cores, baixo poder de processamento, entre outros). Actualmente, as capacidades técnicas são radicalmente diferentes e isso reflecte-se no resultado final dos visualizadores desenvolvidos nos últimos anos. De seguida analisaremos alguns visualizadores que consideramos de referência: o *WhiteCap*² e o *G-Force*³, da autoria de Andy O'Meara, e ainda o *Geiss 2*⁴, o *Smoke*⁵ e o *MilkDrop*,⁶ da autoria de Ryan Geiss, actualmente a trabalhar para a NVidia, desenvolvendo *demos* das capacidades dos seus GPUs.

¹ <http://www.afn.org/~cthugha/>

² <http://www.soundspectrum.com/whitecap/>

³ <http://www.soundspectrum.com/g-force/>

⁴ <http://www.nullsoft.com/free/geiss2/>

⁵ <http://www.nullsoft.com/free/smoke/>

⁶ <http://www.nullsoft.com/free/milkdrop/>

WhiteCap - *Andy O'Meara*

Este visualizador é, em termos conceptuais, bastante simples. Consiste numa visualização do espectro sonoro em cascata, onde a visualização de cada momento temporal vai diminuindo progressivamente de intensidade, até se extinguir por completo. Pela observação que efectuamos deste visualizador, este pareceu-nos o único elemento visual dependente do áudio.

O objecto tridimensional base, uma superfície plana, para além de sofrer as deformações relativas ao espectro sonoro, pode sofrer ainda deformações adicionais, de modo a constituir outras formas (esférica, na primeira imagem, e cilíndrica, na segunda). As cores escolhidas são baseadas em mapas de cor pré-definidos, comutados aleatoriamente. Para aumentar a diversidade visual, a posição e orientação da câmara no espaço tridimensional variam ao longo do tempo.

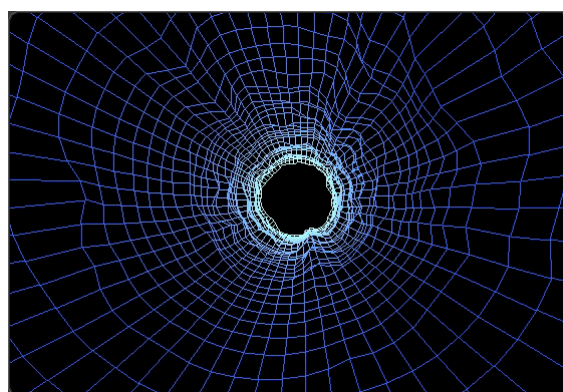
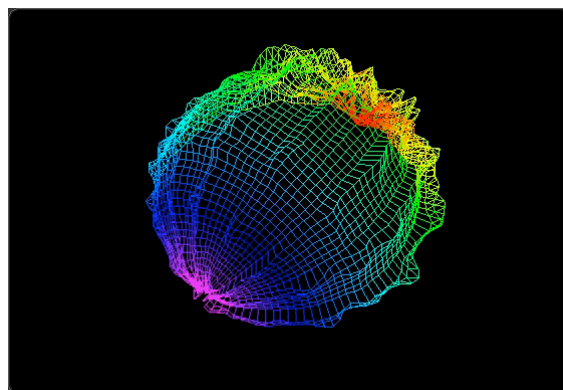


Figura 17 – Visualizador *WhiteCap*

G-Force - *Andy O'Meara*

Esta visualizador é capaz de gerar resultados bastante elaborados. Contrariamente ao anterior, baseado no espectro sonoro, este baseia-se na forma de onda dos sinais como elemento gerador. Estes podem ser distorcidos, como é visível na segunda imagem (aspecto circular). Posteriormente, através de efeitos como o de escala, rotação ou distorção, entre outros, utilizados em conjunto com um sistema de *feedback*, é composta a aparência final. Este visualizador possui uma infinidade de *presets* e as transições suaves entre eles é um dos seus pontos fortes. Mais uma vez, em termos de correlação com o conteúdo musical, não existe muito mais do que a forma de onda original e um mecanismo rudimentar de detecção da pulsação musical (*beat detection*).

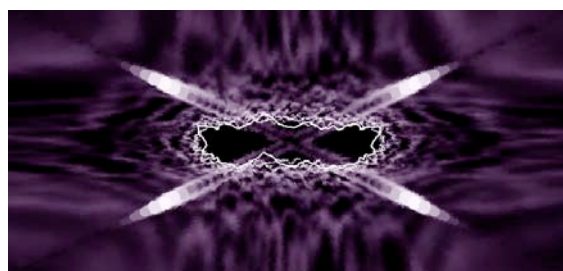


Figura 18 – Visualizador *G-Force*

Geiss 2 - Ryan Geiss

Este visualizador baseia-se também na forma de onda do áudio como elemento gerador. As técnicas utilizadas e os resultados visuais são de certa forma similares ao visualizador anterior, o G-Force. O aspecto que mais nos interessou neste visualizador foi o facto da sua relação com o áudio ser mais efectiva, uma vez que existem mais parâmetros sonoros analisados. Para além da referida forma de onda, possui ainda um sistema de *beat detection* mais elaborado, diferenciando entre *small beat* e *big beat*, e a análise de *noisiness*, ou seja, a maior ou menor quantidade de frequências existentes no sinal.

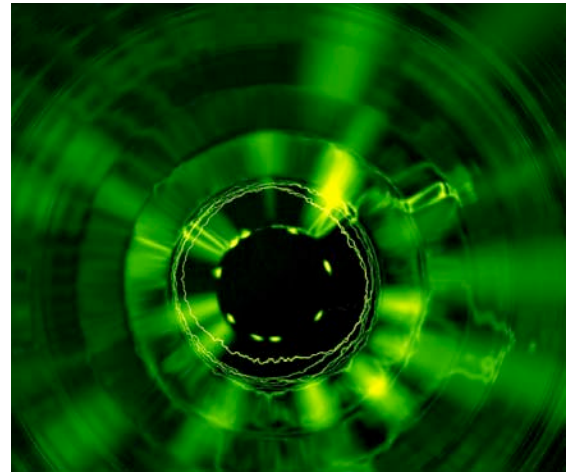


Figura 19 – Visualizador Geiss 2

Smoke - Ryan Geiss

Este visualizador possui uma abordagem diferente de todos os outros, uma vez que não utiliza nem a curva espectral, nem a forma de onda do sinal áudio como base para a criação visual. Neste caso, o elemento gerador consiste em “ondas” controladas pelo áudio através de um sistema de *beat detection*, que se vão deformando através de técnicas de distorção e *feedback* de forma algo aleatória. Ao contrário dos outros visualizadores, este possui apenas um efeito visual, tornando-se por isso algo monótono ao fim de algum tempo. É de referir que este visualizador possui suporte, embora algo rudimentar, para aceleração 3D através do GPU da placa gráfica.



Figura 20 – Visualizador Smoke

MilkDrop – Ryan Geiss

Este visualizador possui um interessante aspecto visual, uma grande variedade de efeitos, e é capaz de obter um elevado número de quadros por segundo (*frame rate*), visto que explora o processamento da placa gráfica (GPU). Ao contrário dos outros visualizadores aqui analisados, este foi concebido de raiz com uma lógica tridimensional, possuindo inclusivamente uma interessante opção de visualização no modo *3D stereo* (com o auxílio de óculos apropriados).

Em termos de correlação com o áudio, este visualizador pareceu-nos muito interessante. O sistema de *beat detection* revelou-se muito preciso, sendo baseado não na intensidade sonora mas na sensação de intensidade, ou seja, o *loudness*, relacionado com a nossa percepção (psicoacústica). Outro aspecto interessante é a activação dos chamados *hard cuts* (mudanças instantâneas de *presets*) através do referido sistema de *beat detection*, sempre que determinado momento musical ultrapassa um determinado nível, resultando numa maior relação entre o áudio e os efeitos visuais gerados.

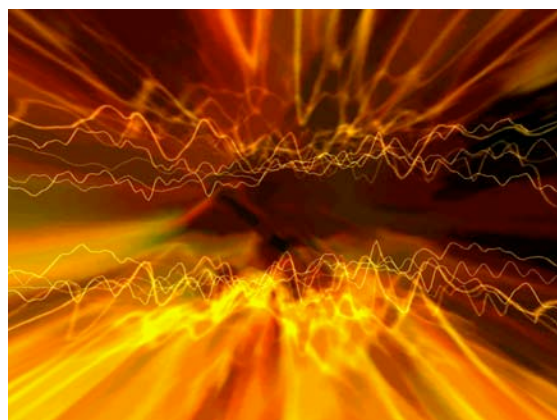


Figura 21 – Visualizador MilkDrop

2.7 Aplicações e ambientes de programação

De seguida abordamos alguns ambientes de programação que nos permitam construir um modelo do *software* idealizado. Para trás ficaram ambientes de programação “tradicionais” baseados apenas em comandos de linha, por três razões fundamentais:

1. Não se adequam ao nosso perfil, sem experiência prévia em linguagens de programação como C, C++, Visual Basic, Java, ou similares;
2. Não se adequam ao aspecto experimental de exploração em tempo-real, durante a fase de programação dos sistemas sonoro e visual;
3. Não possuem uma quantidade mínima de recursos julgados como sendo relevantes. A necessidade de programar algumas funcionalidades básicas imprescindíveis à concretização dos objectivos propostos desviar-nos-ia inevitavelmente dos mesmos.

2.7.1 Análise dos ambientes de programação

Max/MSP e Jitter

O Max da Cycling'74¹ é um ambiente gráfico de programação, ou seja, em vez da tradicional programação através de texto (linhas de código), o programador constrói as aplicações, aqui denominadas de *patches*, através da inclusão de objectos gráficos e da sua interligação através de *patch cords*. Este ambiente de programação é utilizado há já mais de 15 anos por compositores, artistas, professores e alunos, pois proporciona as funcionalidades básicas de uma linguagem de programação (operacionalidade, matemática, lógica, etc) e funcionalidades específicas como a gestão de informação MIDI. Existe a possibilidade de expansão deste ambiente através da programação de objectos adicionais, denominados de *externals*, em linguagens como C ou C++. Estes integram-se neste ambiente de forma transparente, podendo ser utilizados como qualquer outro objecto presente na instalação original. Um ponto forte deste ambiente consiste na possibilidade de compilação do produto final num ficheiro executável (*standalone*).

Sobre esta base, a linguagem de programação Max, foram desenvolvidas duas importantes extensões: o MSP e o Jitter. O pacote MSP² é um conjunto de mais de 150 objectos concebidos para a síntese e *sampling*, a análise e processamento de sinais áudio. Este ambiente permite ainda a inclusão de *plug-ins* VST (*Virtual Sound Technology*) ou a integração com tecnologias como o Rewire³.

¹ <http://www.cycling74.com/>

² <http://www.cycling74.com/products/maxmsp.html>

³ <http://www.propellerheads.se/technologies/rewire/index.cfm?fuseaction=mainframe>

O Jitter¹ expande o ambiente de programação Max/MSP ao proporcionar cerca de 200 objectos com possibilidades de manipulação em tempo-real de imagem, vídeo digital e gráficos tridimensionais, através do suporte para tecnologias como o QuickTime² da Apple e o já referido OpenGL. O Jitter gere toda a informação como matrizes multidimensionais, proporcionando a partilha de dados entre todos os objectos. É possível ainda a comunicação com objectos MSP de forma a transformar áudio em matrizes ou vice-versa. Uma das maiores novidades na versão 1.5, lançada no final de Julho de 2005, é a possibilidade de processamento directamente no GPU (*Graphic Processing Unit*) através da definição de *shaders* programáveis, permitindo o aproveitamento da enorme capacidade de processamento paralelo das actuais placas gráficas.

Dadas estas características, o conjunto Max/MSP/Jitter adequa-se de forma particular às áreas musicais (MIDI, áudio digital) e visuais (imagem e vídeo digital, OpenGL), sendo por isso uma ferramenta com imensas possibilidades de integração multimédia, mais particularmente, no campo audiovisual em que esta tese se situa. O Max/MSP foi inicialmente desenvolvido apenas para o sistema operativo MacOS, e mais tarde para o MacOS X. A versão do Max/MSP para Windows XP foi lançada em 2003, proporcionando uma maior abertura em termos de plataforma. Tanto o Max/MSP como o Jitter são produtos comerciais desenvolvidos pela Cycling'74. Este facto faz-se notar na enorme quantidade de documentação disponível, quer através dos extensos manuais em .pdf, a ajuda directa disponibilizada na própria aplicação, quer nos inúmeros ficheiros de exemplo fornecidos. Este factor possibilita uma curva de aprendizagem mais acessível, particularmente a quem não possuir nenhuma experiência prévia com linguagens e técnicas de programação tradicionais.

Pd e GEM

O Pd consiste também um ambiente de programação gráfica vocacionada para MIDI, áudio, vídeo e gráficos. Tem como raízes o ambiente Max atrás referido, pois o seu autor, Miller Puckette, é responsável por ambos. Um aspecto fundamental no Pd que o diferencia do Max é o seu conceito aberto e gratuito, pois inclui código de muitos programadores, tornando-o num esforço comunitário. Foi construído para ser multi-plataforma, estando disponível para os sistemas Win32, IRIX, GNU/Linux, BSD, e MacOS X. É possível ainda construir *patches* que funcionem também com o ambiente Max/MSP. O objectivo do Pd foi o de “refinar o paradigma do Max com as ideias base de permitir à informação ser tratada de uma forma mais livre e de proporcionar uma abertura a aplicações para lá do áudio e MIDI, como gráficos e vídeo.”³

¹ <http://www.cycling74.com/products/jitter.html>

² <http://www.apple.com/quicktime/>

³ <http://puredata.info/>

O GEM¹ (*Graphics Environment for Multimedia*) funciona, à semelhança do Jitter, como um conjunto de extensões ao ambiente Pd e proporciona a geração em tempo-real de gráficos através da API OpenGL. Foi originalmente desenvolvido por Mark Danks e tem como principal área de aplicação a composição audiovisual. Encontra-se disponível para as plataformas Win32, Linux, IRIX e MacOS X.

AudioMulch

O AudioMulch² é um ambiente de programação similar ao Max/MSP e Pd, permitindo a músicos, compositores e engenheiros de som a criação, análise e processamento de áudio e síntese sonora. Entre as várias potencialidades do AudioMulch, destacamos a sua grande capacidade de utilização em contextos de performance ao vivo. O AudioMulch é um projecto comercial e encontra-se actualmente (versão 2) disponível para as plataformas Windows e MacOS X. Neste momento, esta aplicação apenas possui suporte para o protocolo MIDI, não suportando por isso o protocolo Open Sound Control. No entanto, não deixa de ser um ambiente com potencialidades e algumas características únicas.

Reaktor

O Reaktor é um *software* comercial desenvolvido pela Native Instruments.³ Este não é uma linguagem de programação, mas sim um ambiente gráfico que proporciona a construção de sintetizadores, *samplers*, sequenciadores, processadores de efeitos, mesas de mistura, etc. Pode ser mais frequentemente associado a um instrumento virtual, ou *plug-in*, mas ao contrário da maioria, este é completamente modular.

Esta aplicação, dada a sua versatilidade e facilidade de utilização, é utilizada por inúmeros músicos, engenheiros de som, produtores, e *sound designers*. A facilidade de utilização advém do facto de este permitir a construção de “instrumentos” a vários níveis; desde a junção de instrumentos já concebidos, como de um estúdio virtual se tratasse, até à concepção de cada um deles a partir da combinação de módulos básicos. O perfil dos utilizadores é assim potencialmente bastante diferenciado; desde os utilizadores que apenas utilizam os inúmeros instrumentos já concebidos (pela Native Instruments ou por terceiros) até aqueles que utilizam este ambiente para conceber instrumentos inovadores. Não sendo uma arquitectura aberta, pensada para qualquer tipo de aplicação, mas sim uma aplicação com uma finalidade específica, facilita a construção de instrumentos sonoros, pois garante à partida um conjunto de requisitos importantes como a gestão de polifonia, gestão de *presets* ou memórias, comunicação MIDI e OSC, facilidade de construção de *interfaces* gráficos (painéis de controlo), etc.

¹ <http://gem.iem.at/>

² <http://www.audiomulch.com/>

³ <http://www.native-instruments.com/>

O Reaktor funciona como uma aplicação *stand-alone*, mas também como *plug-in* Audio Unit, VST, RTAS e DXi, o que potencia a sua utilização em ambientes de produção com as chamadas *Digital Audio Workstations* como o ProTools ¹, Logic ², Digital Performer ³, Cubase ⁴, Sonar ⁵, entre outros.

2.7.2 A escolha dos ambientes de programação

A escolha efectuada no que concerne aos ambientes de desenvolvimento do *software* para esta tese de mestrado foi norteada pela relação entre a acessibilidade dos diversos ambientes e as funcionalidades possibilitadas pelos mesmos. Como escolhas finais, optámos pela utilização da aplicação *Reaktor* da *Native Instruments* para a componente sonora do trabalho e a aplicação *Max/MSP/Jitter* da *Cycling'74* para a componente visual. A integração destas duas aplicações é assegurada através do protocolo *OpenSound Control*, utilizando uma ligação de rede local baseada em TCP/IP (protocolo UDP).

Reaktor

Uma escolha equacionada, e igualmente válida seria, a utilização do *Max/MSP/Jitter* para as duas componentes do trabalho (som e imagem). No entanto, a escolha do *Reaktor* para a componente sonora está relacionada essencialmente com o carácter prático, de aplicabilidade a situações reais da indústria que esta tese de mestrado possui. Embora sejam inquestionáveis as capacidades do *Max/MSP* neste domínio, o *Reaktor* pode ser considerado uma ferramenta de mais alto nível, possuindo já muitos elementos que potenciam uma construção mais rápida e intuitiva, sem que isso afecte a qualidade dos resultados. Esta acessibilidade faz com esta aplicação seja uma ferramenta muito utilizada por músicos provenientes de variados géneros musicais, quer em gravações, quer em actuações ao vivo, enquanto que o *Max/MSP*, dadas as suas imensas possibilidades ao nível da programação, é utilizado sobretudo em áreas mais específicas, como a música electro-acústica e experimental, onde as possibilidades de exploração de algoritmos matemáticos para composição e criação sonora possuem grande relevância. Um outro aspecto que nos fez escolher o Reaktor foi a facilidade de criação do *interface* gráfico (Figura 22), comparativamente ao Max/MSP. Visto que o *front-end* do sistema passa pelo controlador MIDI e por este mesmo *interface*, a escolha pareceu-nos acertada.

¹ <http://www.digidesign.com/>

² <http://www.apple.com/logicpro/>

³ <http://www.motu.com/products/software/dp/>

⁴ http://www.steinberg.net/en/products/musicproduction/cubase5_product.html

⁵ <http://www.cakewalk.com/Products/SONAR/>



Figura 22 – Aparência dos dois modos de funcionamento do Reaktor: Estrutura e Painel

Max/MSP/Jitter

Para a componente visual do trabalho, a escolha foi óbvia, pois o ambiente de programação gráfica *Max/MSP* (Figura 23) com a inclusão das extensões *Jitter* não possui neste momento nenhum outro ambiente que se possa equiparar, tendo em conta a relação atrás referida entre acessibilidade e funcionalidade. A documentação deste ambiente é excelente, contendo extensivos manuais de referência, tutoriais, e ficheiros de ajuda integrados na aplicação, para além de inúmeros *patches* que exemplificam uma grande quantidade de operações. O conjunto Pd/GEM fica neste domínio claramente aquém do Max/MSP/Jitter. Quanto às funcionalidades proporcionadas por cada um destes ambientes, a escolha Max/MSP/Jitter pareceu-nos também mais acertada. A extensiva integração do Jitter com tecnologias importantes como a arquitectura multimédia QuickTime, a API OpenGL (agora com suporte para *shaders* programáveis no GPU), a integração de matrizes com estas duas tecnologias e com o resto do ambiente de programação Max/MSP, fizeram-nos escolher este ambiente em vez do Pd/GEM, com capacidades similares no que concerne ao OpenGL, mas inferiores no que concerne ao processamento de vídeo e integração com o QuickTime.

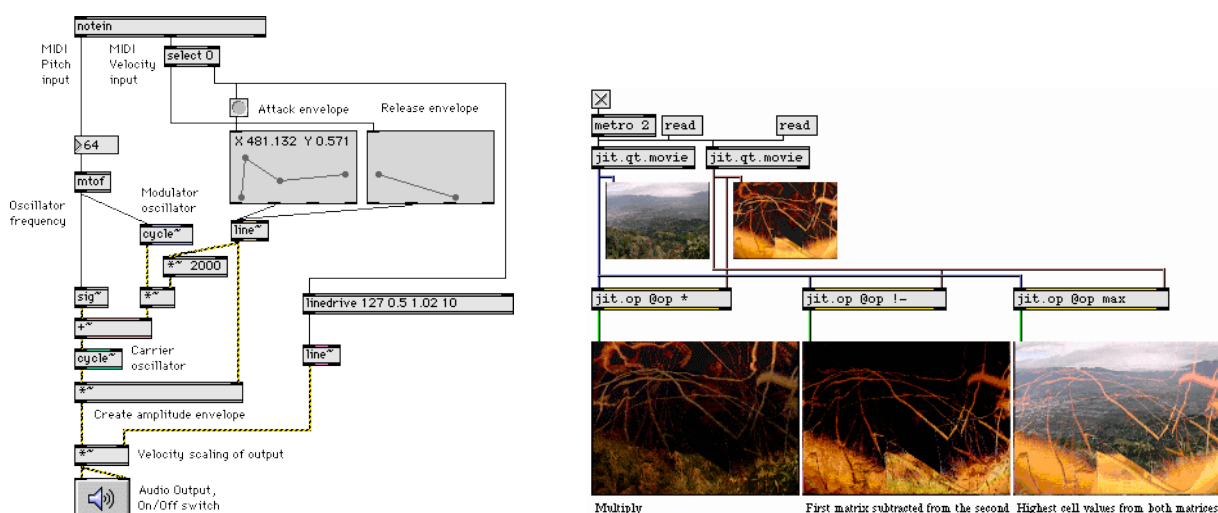


Figura 23 – Aparência do ambiente gráfico de programação Max/MSP (esq.) e Jitter (dir.)

Capítulo III - Implementação

3.1 Introdução

Depois de no capítulo anterior termos abordado um conjunto de oportunidades tecnológicas capazes de potenciar a concepção de instrumentos audiovisuais, é nosso intuito neste capítulo sugerir formas de os concretizar. Uma vez que o sistema proposto tem como premissa fundamental o controlo a um nível musical, sendo a imagem uma consequência do som gerado, torna-se lógico considerar numa primeira fase as possibilidades tecnológicas de controlo do sistema musical e suas respectivas potencialidades. Duas das tecnologias referenciadas anteriormente, o áudio digital e o MIDI, servem de base para a reflexão que se segue (Figura 24).

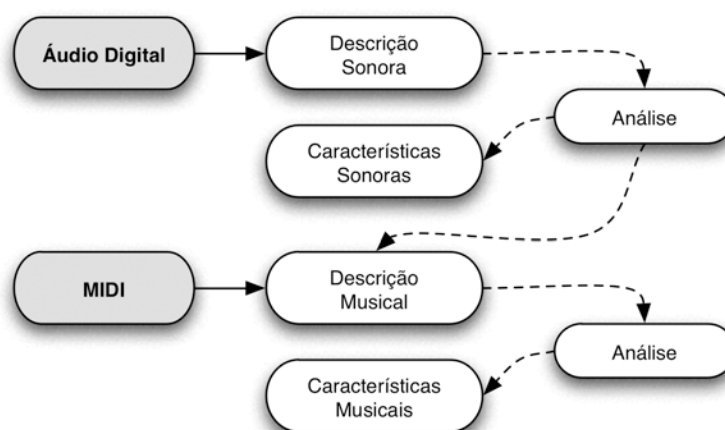


Figura 24 – Diferenças conceituais entre as abordagens partindo do Áudio Digital e do MIDI

O áudio digital, como referido anteriormente, descreve a variação de pressão acústica responsável pelo fenómeno sonoro. Pela observação casual de um tradicional gráfico amplitude/tempo, pouca informação útil conseguiremos extrair para além da chamada “envolvente sonora”, de eventuais momentos de ataque de notas ou da complexidade do sinal. Só através de outros mecanismos de análise, como a conhecida transformada de Fourier, que nos permite uma análise do áudio no domínio frequencial, será possível obter mais informações relevantes acerca do sinal, tais como a sua configuração espectral, a sua harmonicidade, ou a sua centróide. Todas estas características enquadram-se na linha dos descritores de baixo nível definidos pela norma MPEG-7 (ver MPEG-7, pág. 42). No entanto, para o efeito pretendido, a situação mais desejável seria a de podermos obter um reconhecimento sonoro a um nível semântico mais elevado. No caso particular desta tese, esse significado está relacionado com a linguagem musical, logo, serão necessárias técnicas de análise mais evoluídas (descritores de alto nível – MPEG-7), obtidas por vezes através da análise conjunta de várias das técnicas mais básicas atrás referidas. Algumas informações importantes para o objectivo em causa poderão ser o reconhecimento das notas musicais presentes no sinal em cada momento (sons monofónicos ou polifónicos), a intensidade de cada uma dessas notas, a detecção de efeitos como o *vibrato*, o *portamento* ou o *tremolo* ou a detecção do andamento musical. Este é exactamente o tipo de informação descrita pelo protocolo MIDI que, tal como uma partitura, descreve a *performance* do instrumentista e não o som

resultante, pelo que toda a informação musical encontra-se já caracterizada e descrita objectivamente. Assim sendo, quando possível, a utilização do protocolo MIDI elimina a necessidade de análise do sinal, como acontece inevitavelmente no caso do áudio digital, poupando assim recursos de processamento e tornando-se uma solução significativamente mais precisa.

Por fim, um aspecto muito importante que gostaríamos de salientar é o de teoricamente ser ainda possível obter informações resultantes de uma análise musical, tais como o discurso melódico e harmónico, a detecção de motivos melódicos e/ou rítmicos, a forma musical, entre outras. Todas estas informações encontram-se a um nível semântico ainda mais alto e conseguir obter de forma automática esta caracterização seria, sem dúvida, uma situação igualmente desejada para uma correlação audiovisual mais efectiva. Dada a complexidade deste assunto e visto que esta área de conhecimento se distancia já um pouco do assunto que nos propomos abordar, não será desenvolvida nesta tese. No entanto, parece-nos ser uma área com enorme potencial neste contexto de visualização sonora, pelas possibilidades que confere à correlação entre música e imagem a um nível semântico ainda mais elevado.

O relacionamento entre o controlo MIDI e os sintetizadores sonoro e visual poderá ser efectuado de duas formas distintas, representadas nos esquemas da Figura 25. No esquema “A” o MIDI é utilizado para controlar directamente cada um dos sintetizadores. Esta abordagem tem sido a mais utilizada para o sincronismo de conteúdo audiovisual, nomeadamente em equipamentos dedicados e programas de computador orientados para as áreas de VJ,¹ como é o caso dos populares Arkaos VJ,² Resolume,³ Modul8,⁴ VDMX,⁵ ou Zuma.⁶

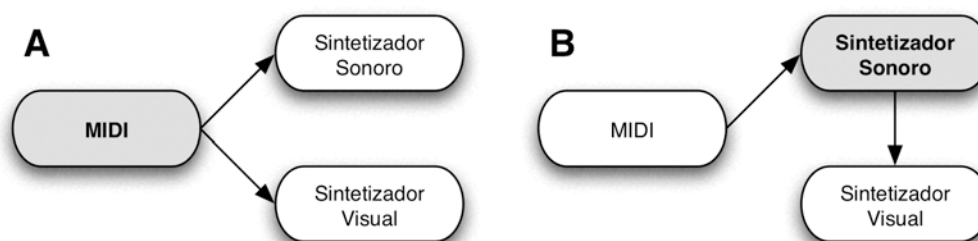


Figura 25 – Possibilidades de relacionamento entre o controlo MIDI e os sintetizadores sonoro e visual

¹ DJ – Disc Jockey / VJ – Video Jockey

² <http://www.arkaos.net/>

³ <http://www.resolume.com/>

⁴ <http://www.modul8.ch/>

⁵ <http://www.vidvox.net/>

⁶ <http://www.3dmaxmedia.com/>

Esta abordagem não permite, contudo, que os inúmeros parâmetros existentes no sintetizador sonoro possam ser partilhados com o sintetizador visual. De facto, uma desvantagem da utilização do protocolo MIDI no actual contexto, é o de este não descrever as características tímbricas do som gerado. A modulação de parâmetros como a amplitude, a afinação ou a frequência de corte de determinado filtro através de envolventes e de LFOs, a configuração de unidades de efeitos e de arpegiadores, todas estas informações não chegam ao sintetizador visual uma vez que não são originárias do MIDI mas sim da configuração interna, ou seja, da programação do sintetizador sonoro. Assim sendo, consideramos que a abordagem ideal em termos de acessibilidade de parâmetros será a resumida no esquema “B” (Figura 25), onde o sintetizador visual é “alimentado” directamente pelo sintetizador sonoro; este fornece-lhe toda a informação necessária, muito mais rica em termos de descrição sonora do que os comandos MIDI que a originaram, pois pode englobar tanto a descrição da *performance* musical como também as características tímbricas do instrumento. Somos da opinião que estas revelam-se como uma mais valia na correlação audiovisual pretendida, pois a função que desempenham no resultado sonoro, particularmente no contexto da música electrónica (onde a exploração tímbrica possui um papel extremamente importante) torna relevante o seu reflexo em termos visuais.

Para que possamos utilizar o sintetizador sonoro como fonte de informação para o sintetizador visual, é necessário que este contemple a possibilidade de a fornecer a outro dispositivo ou aplicação informática. Enquanto que um tradicional sintetizador (módulo de sons) em *hardware* costuma ter apenas saídas de áudio e saída(s) MIDI (normalmente utilizadas para o *backup* dos *presets* através de mensagens de Sistema Exclusivo), as possibilidades tecnológicas que alguns sintetizadores em *software* proporcionam revelam-se neste contexto bastante interessantes, pois ao assentarem numa arquitectura mais aberta, permitem que qualquer parâmetro interno da estrutura do sintetizador, utilizando uma diversidade de tecnologias, possa ser enviado para outras aplicações e até mesmo para outros computadores ou equipamentos dedicados. Programas como o Reaktor, o Max/MSP ou o Pd ¹ possuem este nível de acessibilidade a parâmetros internos. Além do protocolo MIDI, suportam ainda a comunicação através do protocolo OSC, referido no capítulo anterior e que preenche os requisitos necessários para o sistema por nós proposto.

Nesta tese, optámos por abordar o assunto da visualização sonora utilizando estas três “matérias-primas”: o áudio digital, o MIDI e a síntese sonora. Pelo resumo apresentado na Figura 26, para uma tarefa como a que nos propomos realizar, parece-nos lógico que a síntese sonora seja a escolha preferencial, aquela com mais potencialidades em termos de resultados finais, uma vez que todos os parâmetros responsáveis pela produção sonora se encontram já disponíveis. Partindo desta convicção de que a abordagem “sintética” nos oferece melhores condições para a concretização do sistema desejado, decidimos concentrar mais os nossos esforços nesta última. Não pudemos no entanto deixar de abordar de forma menos intensa a visualização de música tendo por base o áudio digital e o protocolo MIDI. A abordagem a estas três tecnologias permitirá retirar

¹ referidos no capítulo 2.7, “Aplicações e ambientes de programação,” pág. 70

algumas conclusões relativamente à viabilidade, possibilidades e requisitos de implementação de cada uma. Todas se revelam úteis na medida em que a aplicabilidade do método de síntese sonora não é de forma alguma universal, bem pelo contrário; terá aplicação prática apenas num determinado contexto de *performance* musical, sendo por isso necessário recorrer também ao áudio digital e ao MIDI como alternativas.

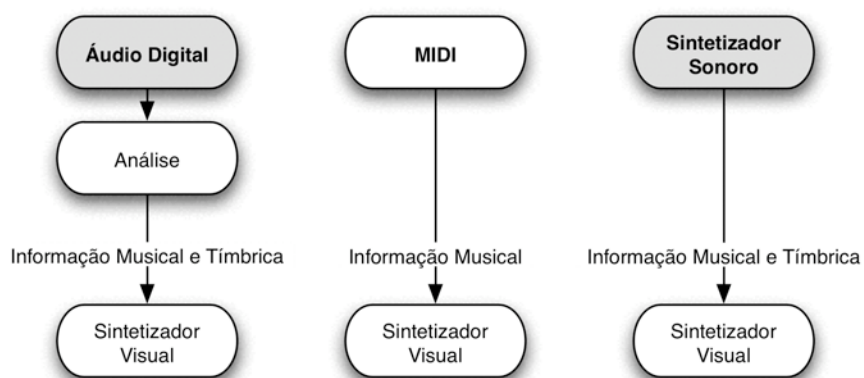


Figura 26 – Diferenças de abordagem entre o Áudio digital, o MIDI e a Síntese sonora

3.2 Comunicação entre aplicações e Processamento distribuído

Tendo em conta a recorrente necessidade de utilizar de forma conjunta diversas aplicações informáticas para alcançar um determinado propósito, dadas as funcionalidades particulares de cada uma, parece-nos relevante explorar as possibilidades actualmente existentes de comunicação entre aplicações, a correr no mesmo computador ou em computadores distintos ligados em rede. Como exemplo desta necessidade, a possibilidade de comunicação revela-se fundamental no contexto desta tese uma vez que o ambiente eleito para a programação do sistema visual (Max/MSP/Jitter) não é o utilizado para a síntese sonora (Reaktor).

Num contexto algo relacionado, de forma a acomodar as eventuais necessidades crescentes do sistema proposto (quantidade de instrumentos musicais), torna-se relevante também abordar possibilidades de implementação de processamento distribuído, quer ao nível do sistema sonoro, quer ao nível do sistema visual. Neste último caso, torna-se essencial rentabilizar, sempre que possível, a enorme capacidade de processamento dos actuais processadores presentes nas placas gráficas (GPU).

3.2.1 Transmissão de MIDI e Áudio entre aplicações locais e em rede

Pretendemos aqui descrever, de uma forma sucinta, a exploração por nós efectuada no decorrer desta tese no que concerne às possibilidades existentes de comunicação de informação MIDI e de áudio entre aplicações, quer a um nível local, quer em rede. Outra tecnologia referida anteriormente, o OSC, dada a sua natureza (protocolo ao nível da aplicação transportado através de UDP), garante já de forma natural estas possibilidades. A necessidade deste tipo de funcionalidades resulta de dois motivos:

- necessidade de utilização e de interligação de diversas ferramentas pré-existentes, como sequenciadores MIDI/Áudio, instrumentos virtuais e leitores de áudio;
- necessidade de comunicação entre computadores, resultantes da abordagem a sistemas de processamento distribuído (sistema sonoro + sistema visual).

Nesta abordagem apenas referimos tecnologias que existam nos sistemas operativos ou que nele se integrem, e que proporcionem funcionalidades a qualquer aplicação que necessite de recursos MIDI e/ou áudio. Deixámos de lado a referência a protocolos proprietários como o Rewire¹ (Propellerhead Software), pois estes necessitam de suporte directo por parte das aplicações para poderem funcionar.

Gostaríamos de sublinhar que, curiosamente, é ainda mais complexo do que seria desejável (especialmente em ambientes Windows) obter este tipo de funcionalidades que deveríamos ter já como garantidas. Em nossa

¹ <http://www.propellerheads.se/technologies/rewire/index.cfm?fuseaction=mainframe>

opinião, o problema constatado só pode ser resolvido pelos sistemas operativos, pois é necessária a gestão transparente e flexível do fluxo de informação áudio e MIDI entre todas as aplicações. A este nível, os vários sistemas operativos baseados no *kernel* do Linux possuem acesso a uma infra-estrutura pensada de raiz para responder a este tipo de utilização: o JACK Audio Connection Kit.¹ Abordamos de seguida apenas as soluções nos ambientes Mac OS X e Windows visto que as aplicações eleitas para o desenvolvimento do *software* proposto (Max/MSP/Jitter e Reaktor) apenas existem nestas plataformas.

Com a crescente necessidade de poder de processamento por parte de músicos e engenheiros de som, resultado da convergência de diferentes funcionalidades no computador (gravador, mesa de mistura, processador de efeitos, sintetizador, *sampler*, etc.), é hoje prática comum a utilização de vários computadores em simultâneo para a execução de diferentes funções. Neste novo contexto, as funcionalidades aqui abordadas tornam-se ainda mais relevantes.

MIDI

Em computadores Apple, a solução para estes problemas encontra-se assegurada através da utilização da tecnologia CoreAudio² já integrada no próprio sistema operativo MacOS X. Esta tem como função gerir todo o fluxo de informação Áudio e MIDI do sistema operativo. Em termos de funcionalidades MIDI permite, entre outras coisas, a partilha de informação entre aplicações residentes no mesmo computador (*IAC Driver*), ou entre aplicações residentes em diferentes computadores, interligados por rede local ou Internet (*Network Driver*). O facto desta tecnologia ser parte integrante do próprio sistema operativo possui vantagens óbvias em termos de facilidade de utilização, fiabilidade e suporte por parte de todas as aplicações.

Em ambientes Windows, embora este tipo de funcionalidades não esteja contemplado pelo sistema operativo, existem algumas soluções que permitem ultrapassar o problema. Uma delas denomina-se “MIDI Yoke”³ e pode ser considerado como um “*driver* virtual” que possibilita a partilha de informação MIDI entre aplicações. Esta solução possibilita apenas a comunicação entre aplicações que estejam a correr no próprio computador. Uma solução mais interessante, que possibilita a comunicação a um nível local ou de rede (como acontece com o MacOS X), é a aplicação “MIDIoverLAN”.⁴ Para além destas funcionalidades, tem ainda a vantagem de possibilitar a comunicação entre as plataformas MacOS X e Windows, visto que se encontra disponível para ambas.

¹ <http://jackaudio.org/>

² <http://www.apple.com/macosx/features/coreaudio/>

³ <http://www.midiox.com/index.htm>

⁴ http://www.musiclab.com/products/rpl_info.htm

Áudio Digital

Em ambientes MacOS X, a tecnologia CoreAudio, ao controlar totalmente o fluxo de áudio e MIDI de todas as aplicações que correm no sistema operativo, pode ser considerada como um “servidor de áudio”. Visto que nenhuma aplicação comunica directamente com as placas de áudio, mas através do CoreAudio, será simples a construção de um *interface* que disponibilize uma espécie de “*patch-bay* virtual” que permita o endereçamento de *streams* de áudio provenientes de uma determinada aplicação para outra. Infelizmente, apesar de todas as infra-estruturas estarem criadas, este último componente não se encontra ainda presente na actual versão do sistema operativo (MacOS X, versão 10.6 – Snow Leopard). Apesar desta limitação, o MacOS X possui já este tipo de funcionalidade disponibilizada sob a forma de *plug-in* Audio Unit (AU): o AUNetSend e o AUNetReceive. Estes podem ser utilizados em programas de sequenciação áudio e MIDI como o Logic e o Digital Performer, entre outros, para enviar e receber *streams* de áudio a nível local ou até mesmo em rede.

Existe no entanto *software* gratuito adicional que permite a funcionalidade pretendida para qualquer aplicação: o “Jack OS X”¹ e o “SoundFlower”². O primeiro é uma implementação para MacOS X do *JACK Audio Connection Kit*, originalmente desenvolvido para o sistema operativo Linux. O Jack OS X é constituído por quatro componentes de software: o servidor, o *driver* virtual, *plug-ins* (AU e VST) e a aplicação *JackPilot*, onde é possível configurar o *patch-bay* virtual. Para necessidades de comunicação de áudio a nível local, com requisitos menos complexos, poderá também ser utilizado o *SoundFlower*, da Cycling’74. Este consiste apenas num *driver* de áudio (placa de áudio virtual) de 2 ou 16 canais que disponibiliza os sinais de saída nas suas entradas.

Em ambientes Windows, torna-se mais complicada a comunicação de áudio entre aplicações. Existem várias soluções ao nível do utilizador comum que permitem muito simplesmente captar o áudio de determinada aplicação. Uma destas é proporcionada pelo *software* de algumas placas de áudio, que mediante a selecção de uma entrada virtual (ex: *wave*), permitem “alimentar” a entrada de áudio com o conteúdo da própria saída. Como exemplos práticos desta necessidade podemos referir a gravação áudio de uma conversa num programa de *audio chat* ou *video chat* ou de um *stream* áudio de determinado *website* ou jogo de computador. Para além desta funcionalidade dos *drivers* de algumas placas de áudio, existem ainda pequenas aplicações que, ao que nos foi dado apurar, não se enquadram convenientemente num contexto mais exigente, tanto em termos de capacidades como também de fiabilidade. Uma excepção encontrada é a aplicação “Virtual Audio Cable” (VAC),³ que funciona de forma algo semelhante ao “SoundFlower”. No entanto, este *software* não possui *drivers* ASIO, apenas *Direct Sound*. Em termos práticos, este facto traduz-

¹ <http://www.jackosx.com/>

² <http://www.cycling74.com/products/soundflower.html>

³ <http://software.muzychenko.net/eng/vac.html>

se em valores de latência demasiado altos para uma utilização exigente em tempo-real, inviabilizando esta solução para aplicações profissionais com necessidades muito particulares ao nível temporal.

3.2.2 Processamento distribuído

Apesar das crescentes capacidades dos computadores ao nível do processamento, seja através do aumento da frequência de relógio, da adição de processadores (ex: *Dual ou Quad Processing*) ou da própria arquitectura dos mesmos (ex: *Dual Core*, 64-bit), a realidade é que também nós, enquanto utilizadores dessa mesma tecnologia, exigimos mais dos equipamentos, passando a utilizar os mesmos em contextos até então não considerados. Daí que, por muito capaz que seja o poder de processamento dos computadores, por vezes tenhamos que recorrer a técnicas de processamento distribuído. O sistema proposto nesta tese, devido à sua potencial complexidade, deverá ser o mais flexível possível, contemplando a necessidade de distribuir o processamento por vários computadores.

A comunicação entre os vários elementos da rede informática é assegurada através de qualquer tecnologia de rede que suporte a arquitectura protocolar TCP/IP (redes locais sem fios, *Ethernet*, *Firewire*, etc). Uma vez que esta arquitectura é actualmente a mais implementada em termos de indústria, este facto confere uma grande abertura ao sistema proposto, permitindo a este funcionar até através da própria *Internet*. Através da tecnologia de rede utilizada poderão ser transmitidas informações OSC, MIDI ou mesmo áudio digital. No decorrer da implementação do sistema proposto, utilizámos a comum tecnologia *Ethernet* a 100 Mbps para a interligação de computadores. Uma vez que as necessidades de transferência do sistema em termos de débito binário não são consideráveis, esta revelou-se extremamente fiável. É conveniente referir ainda que, funcionando o sistema em tempo-real, ao escolher a tecnologia de rede a ser implementada será fundamental considerar também o atraso (latência) e a variação do atraso (*jitter*) típicos da mesma, parâmetros tão importantes neste contexto como o próprio débito binário [63].

Na primeira abordagem efectuada, onde partimos do áudio digital para a representação visual da matéria sonora, as necessidades de processamento do sistema podem requerer apenas a utilização de um único computador que analise o áudio através de várias técnicas e simultaneamente seja responsável pela geração do conteúdo visual. No entanto, caso as técnicas de análise e extracção de parâmetros, dada a sua complexidade, requeiram uma maior capacidade de processamento, poderemos eventualmente dedicar um outro computador para essa tarefa, libertando o anterior para o sintetizador visual. Dada a natureza dos métodos de análise de áudio implementados nesta tese, o recurso ao processamento distribuído não se revelou necessário.

A abordagem acima referida enquadra-se na análise de áudio em *mono* ou *stereo*, contendo já uma mistura de todos os instrumentos e vozes intervenientes. No contexto de uma *performance* musical em concerto, a informação áudio relativa a cada interveniente encontra-se normalmente acessível de forma independente, ou

seja, em canais áudio separados. Nesta situação, as necessidades de processamento serão seguramente muito superiores, pelo que se poderão utilizar esquemas similares aos abaixo propostos.

Nas abordagens que partem do MIDI e da síntese sonora como fonte de informação para o sintetizador visual, o sistema pode tornar-se consideravelmente mais exigente do que na situação anterior. Neste caso, o processamento distribuído possuirá um papel bastante mais significativo. Pensando num contexto de *performance* em tempo-real, como um concerto de música electrónica, por uma questão de fiabilidade consideramos ser essencial que o músico possua um computador dedicado para a síntese sonora (ou sintetizador baseado em *hardware*), à partida localizado no palco, junto de si, uma vez que necessitará de interagir com o mesmo durante o concerto. Este computador pode estar ligado através de uma tecnologia de rede ao computador responsável pela síntese visual, localizado junto do sistema de projecção de vídeo, do técnico de luz, de um VJ, ou de qualquer outro elemento.

Uma outra abordagem possível em termos de processamento distribuído, será a de vários músicos, cada um com o seu computador ou sintetizador dedicado colaborarem em simultâneo quer para o ambiente sonoro, quer para os efeitos visuais gerados. Esta possibilidade revela-se extremamente interessante em termos estéticos, uma vez que a função de cada músico/instrumento no contexto musical poderá estar também representada em termos visuais de formas completamente distintas.

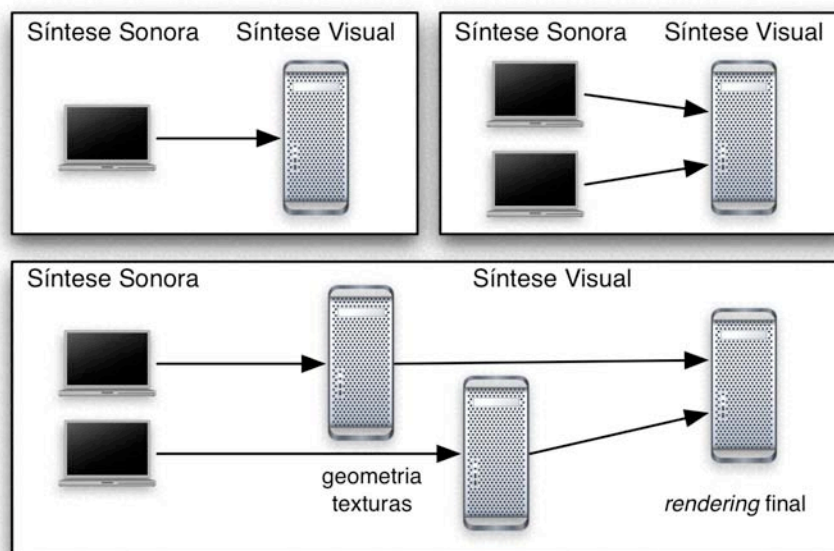


Figura 27 – Possibilidades de processamento distribuído

Por último, numa situação de processamento já consideravelmente mais complexa (Figura 27), é considerada ainda a possibilidade de cada um dos “instrumentos” possuir um computador dedicado para a geração da sua representação visual. Cada um destes computadores poderá efectuar todos os cálculos relativos à definição de geometria e texturas *bitmap*, dados esses que são transmitidos a um computador responsável por aplicar as

texturas nos objectos e representá-los no espaço visual, definir as possibilidades de interacção entre eles e executar o *rendering* final da cena tridimensional. O ambiente de programação utilizado para concretizar a implementação do sintetizador visual, o Max/MSP/Jitter, possui a capacidade de comunicar este tipo de informações por rede, através dos objectos [jit.net.send] e [jit.net.receive]. Estes podem enviar e receber matrizes, tanto de definição de geometria como de texturas, através da arquitectura protocolar TCP/IP.

É conveniente referir que se a programação do sistema visual implementar a maior parte do processamento no GPU em detrimento do CPU do computador, a necessidade de recorrer a estes sistemas de processamento distribuído torna-se consideravelmente menor, uma vez que a capacidade de processamento das actuais placas gráficas é já extremamente elevada.

CPU vs GPU

A utilização de um processador dedicado para a geração e processamento de gráficos – o *GPU* – pode ser considerada já, pelo menos em termos conceptuais, como uma forma de aliviar o processamento do processador principal. Tradicionalmente, todo o processamento gráfico ao nível de texturas (*fragment*) e de vértices (*vertex*) tem vindo a ser calculado pelo processador principal do computador (CPU). No entanto, e visto que a *performance* e funcionalidade dos processadores incluídos nas placas gráficas tem vindo a aumentar consideravelmente, temos assistido a uma crescente utilização destes últimos para o processamento atrás mencionado. Esta possibilidade só é concretizável se a placa gráfica possuir suporte para aquilo a que designamos de *shaders* programáveis.¹ Através destes programas, é possível instruir o GPU a executar determinado tipo de processamento não incluído na *pipeline* fixa do OpenGL. A utilização destes *shaders* possibilita um aumento considerável no desempenho do sistema, uma vez que o processamento se efectua ao nível do GPU. Como contrapartida, ao ser um processador dedicado, não permite tanta flexibilidade como um processador genérico, logo poderemos ficar algo limitados quanto ao tipo de algoritmos a executar neste.

Durante o período de realização desta tese, a possibilidade de acesso a estas tecnologias aumentou consideravelmente, pelo que nos sentimos impelidos a tirar partido das mesmas nos protótipos realizados. Conseguimos assim passar progressivamente de uma implementação totalmente baseada em processamento efectuado pelo CPU para uma outra em que a maior parte deste é feito pelo GPU através da tecnologia de *shaders* programáveis.² É conveniente notar que com este tipo de implementação conseguimos obter uma *performance* do sistema muitas vezes superior à implementação original,³ quando consideradas a resolução e o número de imagens geradas por segundo (*framerate*). Tendo em conta a tendência crescente de utilização

¹ consultar a secção “Shaders”, pág. 63

² Dado não possuímos qualquer experiência prévia no desenvolvimento de *shaders* nem a programação destes se enquadra no intuito da presente tese, decidimos limitar as nossas possibilidades de trabalho a *shaders* previamente programados. Desta forma, ficamos um pouco restringidos no tipo de processamento disponível ao nível do GPU.

³ dependendo da placa gráfica e processador utilizados.

de formatos de vídeo de alta definição e a necessidade de um *framerate* elevado, quer por motivos de fluidez de movimento, quer pela rapidez da resposta visual ao impulso sonoro, torna-se absolutamente desejável potenciar a utilização do GPU.

Quando o processamento pretendido não se revelar viável de ser calculado pelo GPU, quer por impossibilidades técnicas ou por incapacidade na programação de novos *shaders*, torna-se necessária a utilização do CPU. Para tal, as texturas (imagens *bitmap*) armazenadas na VRAM,¹ através de um processo denominado de *texture readback* são enviadas para a memória principal do computador (RAM). Depois da textura ser processada no CPU poderá voltar novamente ao GPU para posterior processamento. Este procedimento acarreta normalmente uma perda de desempenho significativa (decréscimo do *frame rate* em função da resolução das texturas utilizadas), pelo que só deverá ser efectuado em último recurso. Para que este processo não comprometa ainda mais o desempenho do sistema, é aconselhável a utilização de *slots* de comunicação *PCI-Express* (x16) em detrimento dos anteriores *AGP*. Isto porque esta tecnologia permite uma comunicação entre o computador (CPU, RAM) e a placa gráfica muito mais eficiente, com uma largura de banda de 4 GB/s em cada um dos sentidos em simultâneo (a tecnologia *AGP* (8x) permite apenas 2,1 GB/s no total).

¹ VRAM = Video RAM

3.3 Visualização de Áudio Digital

Nesta primeira abordagem, optámos por utilizar como fonte de informação o áudio digital, uma vez que esta é provavelmente a fonte mais utilizada e com um alcance mais abrangente, como pode ser comprovado pela incorporação de módulos de visualização em programas como o iTunes¹, Windows Media Player² ou Winamp³. Esta experiência serve assim como termo de comparação para os próximos métodos abordados, onde utilizamos o MIDI e a síntese sonora como ponto de partida para a exploração visual.

De uma forma geral, os chamados “visualizadores de áudio” analisam o áudio através de métodos simples, extraindo um conjunto de parâmetros que servirão de base para a correlação com algum aspecto visual. É de salientar a necessidade desta análise ser efectuada em tempo-real dado o carácter deste tipo de aplicações. Os parâmetros detectados nunca são, de uma forma geral muito complexos. Entre os mais comuns encontra-se a detecção de tempos fortes na música. Este parâmetro funciona melhor para determinados estilos musicais onde a bateria possui um papel importante, visto que esta detecção está normalmente optimizada para a zona de frequências emitidas pelo bombo ou tarola. Outro aspecto que deriva directamente do áudio é a forma de onda do sinal, que funciona muitas vezes como elemento gerador do conteúdo visual.

A visualização de música tendo como fonte de informação o áudio digital pode ser considerada do ponto de vista prático como inferior às restantes abordagens aqui mencionadas por requerer uma etapa de análise que, dada a sua complexidade, nem sempre se revela precisa. No entanto, no contexto desta tese, a recorrência a sinais áudio como matéria-prima para o sistema visual torna-se necessária em duas situações distintas: na visualização de música previamente gravada e misturada, ou em situações de *performance* com instrumentos musicais acústicos/electro-acústicos, grupo em que se insere a maior parte dos instrumentos.

3.3.1 Música pré-gravada

Apesar do enfoque principal da presente tese se direccionar para a área da *performance* em tempo-real, dada a popularidade de abordagens neste contexto, parece-nos relevante a referência de algumas particularidades. A visualização de material sonoro previamente gravado e misturado possui alguns inconvenientes quando comparado com a situação de *performance* em concerto que de seguida abordaremos. Os recursos de análise necessitam de ser bastante mais complexos, uma vez que todo o material se encontra já misturado. Torna-se assim substancialmente mais difícil detectar notas e outras características de um determinado instrumento. Com este tipo de sinais, será mais sensato efectuar uma análise às características sonoras na sua globalidade,

¹ <http://www.apple.com/itunes/overview/>

² <http://www.microsoft.com/windows/windowsmedia/mp10/>

³ <http://www.winamp.com/>

não atentando tanto aos aspectos individuais de cada um dos instrumentos. No entanto, algumas excepções poderão ser feitas; num contexto de géneros musicais populares como a *Dance Music* ou o *Pop-Rock*, elementos como o bombo ou a tarola de uma bateria poderão fornecer informações importantes para determinar o ritmo visual do sistema proposto. Estas informações são possíveis de obter, por um lado pela zona frequencial que estes elementos preferencialmente ocupam e, por outro, por serem sons percussivos, possuidores de um ataque bastante pronunciado.

A utilização de ficheiros de áudio tendo em vista a visualização musical possui contudo uma vantagem relativamente à *performance* musical ao vivo: a possibilidade de descrição e catalogação prévia do seu conteúdo através de *metadata* (como acontece com os conhecidos *ID3 Tags* característicos do formato *mp3*). Este facto revela-se extremamente interessante em termos de visualização, uma vez que permite a utilização de qualquer um desses dados tendo em vista uma mais efectiva correlação audiovisual. Algumas informações eventualmente úteis sobre determinada música poderão ser o género musical em que se insere, o ano da sua publicação ou até o nome do artista. Estas e outras informações poderão accionar um conjunto de parâmetros do sintetizador visual (*presets* pré-configurados), de forma a que este possa fornecer um enquadramento estético apropriado. A título de exemplo, seria possível obter automaticamente imagens a partir da internet (ex: Google Images¹) com base no nome do intérprete (presente na *metadata* do ficheiro) e utilizá-las como texturas no sistema visual idealizado.

3.3.2 Performance ao vivo

Num contexto de *performance* ao vivo, em concerto, cada instrumento é habitualmente captado por um ou até vários microfones de forma individualizada. Assim sendo, toda a informação musical encontra-se na mesa de mistura em canais áudio separados de modo a poderem ser manipulados de forma independente. Nestes casos, a possibilidade que temos em aceder selectivamente a cada um dos instrumentos constitui uma enorme vantagem para a visualização musical por um conjunto de motivos. Por um lado, simplifica consideravelmente os métodos de análise. Visto que um dos mais complexos objectivos pretendidos pela análise de áudio seria o reconhecimento de elementos distintos de uma mistura,² com a separação da matéria-prima sonora em canais de áudio distintos, esta necessidade fica de imediato assegurada. Por outro lado, pode potencialmente tornar os resultados da análise muito mais precisos, uma vez que cada um dos sinais áudio é consideravelmente mais simples do que a mistura final. Desta forma, torna-se perfeitamente viável, por exemplo, efectuar a detecção de uma linha de baixo, de uma voz principal ou de cada uma das peças de uma bateria. Em termos estéticos/artísticos, este procedimento torna possível a visualização selectiva de instrumentos, ou a aplicação de representações visuais distintas para cada um dos intervenientes, segundo a sua relevância no contexto musical.

¹ <http://images.google.com/>

² Em determinados casos, até de instrumentos que partilham parcialmente a mesma zona frequencial.

Para efectuar este tipo de implementação, torna-se necessário que a mesa de mistura possa enviar os sinais pretendidos para o computador responsável pela análise do áudio e que este possua condições para o receber (Figura 28). No primeiro requisito, quase todas as mesas de mistura, quer sejam concebidas para trabalho em estúdio, quer para som ao vivo, costumam possuir saídas individuais por canal com o propósito de enviar os sinais áudio para equipamentos de gravação ou outros, designadas habitualmente por termos como *Direct Out*, *Tape Send* ou *Multi Out*. Poderão alternativamente ser utilizadas as saídas de auxiliar, normalmente associadas ao envio para monição ou processadores de efeitos. Qualquer uma destas situações permite o envio dos sinais pretendidos para o sistema responsável pela análise dos sinais áudio. Em condições ideais, este envio da mesa de mistura deverá encontrar-se configurado para *post-fader*, para que o áudio seja enviado para o computador responsável pela sua análise após passar pelo *fader* do respectivo canal, estágio de ajuste de volume do instrumento por parte do técnico de som. Desta forma, o próprio equilíbrio dos instrumentos na mistura pode ser representado visualmente uma vez que o nível de envio de cada canal é afectado pelas opções técnicas e artísticas do engenheiro de som. Por fim, para que o computador possa receber os vários canais de áudio, torna-se imprescindível a utilização de uma placa de áudio multi-canal adequada quanto ao número de entradas e de protocolos de áudio analógico/digital disponíveis, de forma a se compatibilizar com as necessidades da *performance* em causa e com a mesa de mistura utilizada.

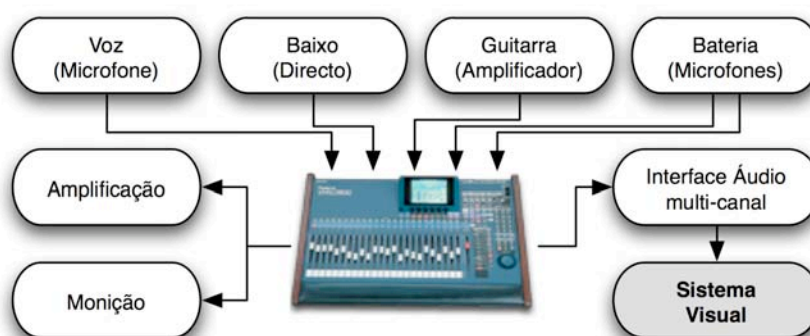


Figura 28 – Exemplo de visualização de áudio em ambiente multi-pista

3.3.3 Recursos de análise disponíveis

Nesta tese não nos propomos desenvolver novas formas de análise do áudio, pois esse assunto tem dado e continuará a dar muitas teses por si só. Assim sendo, decidimos utilizar métodos ou ferramentas desenvolvidas por terceiros que nos proporcionem já alguns parâmetros úteis para a correlação visual. No ambiente de programação por nós utilizado, o Max/MSP/Jitter, existem já objectos que nos proporcionam algumas informações relativamente ao áudio. Passamos a enunciar os mais importantes:

[snapshot~]

Este objecto permite efectuar amostras do sinal. Sempre que este é accionado por um evento, o valor de amplitude do sinal é reportado sob a forma de um *float*.

- [meter~] [peakamp~]** Estes objectos permitem obter o valor máximo de amplitude do sinal.
- [average~] [avg~]** Estes objectos permitem obter o valor médio de um sinal durante um determinado período de tempo.
- [fft~] e [ifft~]** Estes dois objectos permitem efectuar a transformada de Fourier num sinal, ou seja, transformar a representação do sinal de um domínio temporal para um domínio frequencial [fft~] e vice-versa [ifft~]. É com base nesta técnica que a maior parte dos métodos de análise de parâmetros sonoros se baseiam, pois permite, por exemplo, analisar a estrutura harmónica de um sinal, podendo-se retirar daí imensas conclusões através de processamento adicional.
- [jit.catch~]** Neste caso particular, não é feita nenhuma análise ao sinal. Contudo, resolvemos mencionar este objecto pois é responsável pela conversão de um sinal MSP numa matriz Jitter. Possui vários modos de funcionamento relativamente à captura dos sinais, sendo o seu comportamento similar ao de um comum osciloscópio.

Existem outros objectos no Max/MSP, como o [scope~] e o [spectroscope~] que nos proporcionam a análise de sinais. No entanto, para a tarefa que nos propomos desenvolver estes não nos serão úteis, uma vez que não possuem *outlets* capazes de enviar informação discreta para posterior manipulação. São principalmente utilizados para monitorizar sinais no ambiente Max/MSP durante a fase de programação.

Seria possível, à partida, através destes objectos acima mencionados (com destaque particular para o [fft~]), desenvolver métodos de extracção de um conjunto de parâmetros sonoros de maior complexidade, como a informação de altura (afinação) ou dos momentos de ataque de eventos musicais. Felizmente, existem objectos adicionais desenvolvidos para o ambiente Max/MSP por terceiros que nos proporcionam algumas dessas funcionalidades. Todos os objectos referidos funcionam também em tempo-real, requisito fundamental para o nosso objectivo. Aqui ficam alguns dos mais importantes por nós testados.

- [fiddle~]** Objecto que permite decompor o sinal em sinusóides e detectar a altura das notas existentes num sinal. Este objecto foi desenvolvido pelo próprio Miller Puckette¹, e o *port* para MSP é da responsabilidade de Ted Apel e David Zicarelli.
- [bonk~]** Este objecto permite detectar o ataque de instrumentos percussivos. Foi criado também por Miller Puckette e convertido para MSP por Ted Apel. Mais informações sobre o desenvolvimento deste objecto, bem como do [fiddle~] podem ser encontradas num *paper* dos referidos autores intitulado “*Real-time audio analysis tools for Pd and MSP*”².
- [centroid~]** Este objecto desenvolvido por Ted Apel, John Puterbaugh e David Zicarelli permite calcular o centro espectral de um sinal.

¹ Miller Puckette é o autor original do ambiente de programação Max (actualmente desenvolvido e comercializado pela Cycling’74), e mais tarde, também do Pure Data, que desenvolve até hoje segundo a filosofia *open source*.

² <http://www-crcs.ucsd.edu/%7Etapel/icmc98.pdf>

Estes três objectos estão disponíveis para MacOS 9 e X na página de Ted Apel¹. No caso da utilização do Max/MSP para Windows, existem *ports* disponíveis dos objectos [fiddle~] e [bonk~]². Um outro conjunto de objectos extremamente interessantes para a análise de áudio são os criados por Tristan Jehan³ [64], do Instituto de Tecnologia de Massachusetts, para os sistemas MacOS 9 e X, e com conversão para Windows por Paul Hill. Estes objectos proporcionam as seguintes capacidades de análise:

- [pitch~]** Este objecto, inspirado no [fiddle~] acima descrito, serve também para detectar a altura das notas presentes no sinal.
- [loudness~]** Ao contrário dos habituais objectos [average~] ou [peakamp~], este permite obter a sensação de intensidade, relacionada com a nossa percepção fisiológica.
- [brightness~]** Este objecto é similar ao [centroid~], permitindo medir o brilho de um sinal, ou seja, o seu centro espectral.
- [noisiness~]** A quantidade de “ruído” de um sinal pode ser medida através deste objecto, ou seja, a quantidade de frequências distintas presentes no sinal.
- [bark~]** Decomposição espectral segundo a escala de Bark. Analisa o espectro do sinal através de 25 bandas; 0, 100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4400, 5300, 6400, 7700, 9500, 12000, 15500
- [analyzer~]** Este objecto analisa todos os parâmetros acima referidos através da utilização de um só objecto.
- [segment~]** Este objecto detecta a existência de sons de altura definida ou não definida.
- [beat~]** Este objecto permite detectar o andamento e os ataques de uma música.

Para a realização desta componente do trabalho, efectuaram-se testes a estes objectos de forma a extrair parâmetros eventualmente úteis para a correlação audiovisual. Depois de analisada a funcionalidade e fiabilidade destes, foram seleccionados para inclusão neste trabalho os objectos [analyzer~] e [beat~]. O primeiro é extremamente interessante pois permite-nos obter todas as informações relativas aos objectos [pitch~], [loudness~], [brightness~], [noisiness~] e [bark~]. A vantagem inerente à utilização deste objecto em detrimento dos restantes separadamente, para além da maior conveniência que proporciona, é a de poupar recursos de processamento (a operação FFT necessita de ser executada apenas uma vez, enquanto que cada um dos outros objectos teria que o fazer em separado). O objecto [beat~] permite calcular a presença de “ataques” no sinal e o andamento da música, de acordo com a cadência de ataques capturada.

É ainda de referir que depois dos sinais serem discretizados em eventos pelos objectos acima referidos, torna-se possível efectuar outros processos de análise, com os objectos incluídos no Max/MSP, ou de terceiros, como é o caso daqueles que de seguida descrevemos:

¹ <http://crca.ucsd.edu/~tapel/software.html>

² <http://www.akustische-kunst.org/maxmsp/other.html>

³ <http://web.media.mit.edu/%7Etristan/maxmsp.html>

- tl.objects**¹ Conjunto de *externals* para Max/MSP desenvolvidos por Trond Lossius. Destes objectos, podemos destacar o [tl.velocity], [tl.delta] e [tl.delta2], úteis para a detecção de ataques.
- f0.ext**² Conjunto de *externals* para Max/MSP desenvolvidos por Fredrik Olofsson. Destes objectos, podemos destacar o [f0.beats_to_frames] para calcular o número de *frames* de vídeo em relação à pulsação musical, os [f0.range] e [f0.range2] para o cálculo de valores médios e os [f0.smooth] e [f0.smooth2], úteis para suavizar um determinado conjunto de valores.
- Lobjects**³ Conjunto de *externals* para Max/MSP desenvolvidos por Peter Elsea. Estes objectos adicionam funcionalidade ao Max na área da manipulação de listas de valores.

3.3.4 Programação

Nesta secção não pretendemos descrever em pormenor a totalidade da programação do *patch* criado, nomeadamente, o processamento de efeitos através de matrizes Jitter ou os objectos OpenGL. Isto porque os principais aspectos da construção de todo o sistema visual merecerá especial destaque posteriormente. Serão apenas focados os aspectos exclusivos a esta abordagem de mapeamento entre o Áudio digital e o Sintetizador Visual, partindo de sinais áudio provenientes de ficheiros ou das entradas da placa de som.

Entrada de Áudio

De forma a tornar o sistema mais versátil ao nível da fonte sonora, foi programado um *patch* no Max/MSP que permite obter um sinal áudio proveniente das entradas da placa de som (objecto [adc~]) ou de um ficheiro de áudio pré-gravado (objecto [sfplay~], ou o conjunto [jit.qt.movie]/[spigot~]). O primeiro (*analog-digital converter*), permite que qualquer sinal áudio presente na entrada da placa de som possa ser utilizado como base para a análise e posterior visualização. O segundo (*sound file play*), permite reproduzir ficheiros nos formatos AIFF, SD II, NeXT/SUN(.au), WAVE, e Raw Data. De modo a poder utilizar ainda outro tipo de ficheiros, em particular ficheiros com compressão como os populares MP3 (*MPEG Layer 3*) ou AAC (*Advanced Audio Coding*), decidimos utilizar a arquitectura QuickTime, visto que esta suporta, para além dos formatos áudio já referidos, muitos outros menos comuns. Para podermos utilizar esta última fonte de áudio, o QuickTime, necessitamos de dois objectos: o [jit.qt.movie], responsável pela leitura do ficheiro áudio⁴ desejado, e o [spigot~], que reencaminha o sinal áudio do QuickTime para o MSP, de forma a poder ser processado pelos restantes objectos dentro do ambiente Max/MSP/Jitter.

¹ <http://www.bek.no/~lossius/download/>

² <http://www.fredrikolofsson.com/pages/code-max.html>

³ <http://arts.ucsc.edu/EMS/Music/research/Lobjects.readme.html>

⁴ Neste contexto, os ficheiros são de áudio, mas este objecto permite ler qualquer tipo de ficheiro suportado pela arquitectura QuickTime.

Análise de Áudio

Como referido anteriormente, a maior parte das informações extraídas dos sinais áudio foram provenientes do objecto [analyzer~]. Das informações disponibilizadas por este foram particularmente úteis as saídas de *pitch* (para conteúdo monofónico), *loudness*, *brightness*, *noisiness* e *attack*.

O parâmetro *pitch* é utilizado para detectar a altura de sinais monofónicos, com valores entre 0 e 127, similar ao MIDI. Não sendo um parâmetro particularmente útil no contexto da análise de uma mistura completa, pode revelar-se bastante interessante quando aplicado, por exemplo, numa voz ou instrumento solista principal. O parâmetro *loudness* utilizado é expresso em dB (escala logarítmica). Utilizamos o objecto [zmap] para permitir a limitação e escalamento dos valores recebidos em valores de amplitude entre 0 e 1. O facto de relacionarmos directamente uma unidade de escala logarítmica com uma de escala linear sem uma conversão prévia prende-se com o facto de a escala logarítmica ser aproximadamente linear no que concerne à nossa percepção da amplitude sonora.¹ Os parâmetros *brightness* e *noisiness* são-nos fornecidos pelo objecto [analyzer~] numa escala linear entre 0 e 1, enquanto que o *attack* transmite apenas uma mensagem *bang* (*trigger*), não sendo por isso caracterizado por qualquer valor específico relativamente à intensidade. Podemos contudo utilizar esta informação de ataque para accionar a leitura do valor do parâmetro *loudness* nesse instante, obtendo assim uma quantificação do valor de ataque.

Foi desenvolvida ainda a possibilidade de análise do posicionamento estereofónico de sinais obtido a partir da comparação dos valores de *loudness* de cada canal. É de realçar que, após várias experiências de associação a parâmetros visuais, termos acabado por o considerar pouco significativo, pois na maioria dos casos musicais testados, particularmente em música pré-gravada e misturada não existe uma grande amplitude de variação e, por outro lado, a variação existente possui uma frequência demasiado rápida para se tornar útil. Esta constatação é um reflexo do facto dos engenheiros de som responsáveis pela mistura das músicas terem já em atenção o equilíbrio da “imagem estereofónica”, apenas utilizando desequilíbrios momentaneamente como efeito.

Por fim, um tipo de informação obtido a partir do áudio que se revelou essencial para a construção da estética do sistema visual foi a forma de onda do sinal (gráfico amplitude/tempo). O principal objecto responsável por este processo foi o [jit.catch~], que serve basicamente de “ponte” entre o MSP e o Jitter, recebe um ou mais sinais MSP e transforma-os em matrizes Jitter adequadas a uma posterior manipulação para efeitos de visualização. Este objecto possui vários modos de funcionamento relativamente à captura dos sinais, sendo que o modo 3 seleccionado é similar ao comportamento de um comum osciloscópio.

¹ A lei de Weber-Fechner determina que a sensação cresce como o logaritmo do estímulo.

3.4 Visualização de MIDI

A utilização da tecnologia MIDI tendo em vista a visualização musical possui, graças à sua natureza descritiva, vantagens imediatas relativamente ao áudio digital. Toda a informação encontra-se já digitalizada e discretizada em parâmetros apropriados a um eventual mapeamento audiovisual. Num contexto de *performance* musical, passamos a considerar a utilização de dois tipos de instrumentos: os instrumentos musicais electrónicos e instrumentos acústicos/electro-acústicos.

3.4.1 Controladores e Conversores MIDI

O MIDI tem sido amplamente utilizado pelos fabricantes de instrumentos musicais electrónicos; desde a primeira implementação deste protocolo no Prophet 600 da Sequential Circuits (Figura 29), em 1983, que quase todos os instrumentos musicais electrónicos construídos nos últimos 25 anos o têm vindo a incluir. É de realçar também que esta tecnologia encontra-se implementada em diferentes tipos de instrumentos: apesar do convencional teclado ser o mais comum (sintetizadores, pianos digitais, órgãos), existem também controladores de sopro como os modelos WX da Yamaha (WX7, WX11 ou o actual WX5¹), baterias electrónicas, sendo a série V-Drums² da Roland provavelmente a mais popular, violinos, violas, violoncelos e contrabaixos como os da Zeta Music Systems,³ acordeões como o Roland FR-7,⁴ harpas da Kortier,⁵ entre muitos outros. Para além destes sistemas de controlo, quase sempre inspirados em instrumentos acústicos, tem vindo a ser desenvolvido um conjunto considerável de controladores não convencionais, visto ser uma área de pesquisa e experimentação extremamente activa [49] [56].



Figura 29 – Sintetizador Prophet 600 da Sequential Circuits

¹ <http://www.yamahasynth.com/products/wx5/index.html>

² <http://www.roland.com/V-Drums/>

³ <http://www.zetamusic.com/>

⁴ <http://www.roland.com/products/en/FR-7/>

⁵ <http://www.kortier.com/midi36.htm>

Se os músicos utilizarem o tipo de instrumentos anteriormente referidos, a comunicação através do protocolo MIDI encontra-se já assegurada. No entanto, na maioria dos instrumentos musicais acústicos e electroacústicos, este protocolo não se encontra implementado. Felizmente, em muitos destes instrumentos existem formas efectivas de discretizar a *performance* musical em eventos MIDI através de *hardware* adequado. Alguns exemplos deste tipo de equipamento são o Moog PianoBar,¹ que se instala em pianos acústicos, o Roland GI-20,² interface para guitarra acústica ou eléctrica que converte a informação captada por *pick-ups* como o GK-2A ou o GK-3,³ (Figura 30) ou baterias acústicas equipadas com sensores (*triggers*) como os da DDrum⁴ ou o TMC-6⁵ da Roland. Existem ainda outros tipos de conversores, normalmente monofónicos, normalmente referenciados como conversores *pitch-to-MIDI* [56], úteis para a detecção da *performance* musical em vozes ou outros instrumentos. Neste contexto, continuamos a considerar a fonte de informação musical como sendo o MIDI e não o áudio, visto que a etapa de análise é garantida pelo equipamento externo, relacionando-se o nosso sistema directamente com os eventos já discretizados.



Figura 30 – Pickup GK-3 e Conversor MIDI GI-10, da Roland

3.4.2 Mensagens MIDI e possibilidades de controlo

Esteja o protocolo MIDI já incluído, ou implementado por equipamentos adicionais, a informação é descrita através de mensagens MIDI. De seguida, passamos a considerar, de forma muito resumida, possíveis utilizações para as mensagens MIDI neste contexto de visualização.

- As mensagens mais comuns do protocolo MIDI são, sem dúvida, as de *Note On/Note Off*. Este conjunto de mensagens caracteriza as notas tocadas ao nível da intensidade (dinâmica) e respectiva duração. A informação de *pitch bend* pode juntar-se a estas para alterar a afinação da nota tocada.

¹ <http://www.moogmusic.com/>

² <http://www.roland.com/products/en/GI-20/index.html>

³ Em algumas guitarras este *pick-up* vem já incluído, como é o caso de vários modelos da marca Godin.

⁴ <http://ddrum.com/drumtriggers.htm>

⁵ <http://www.roland.com/products/en/TMC-6/index.html>

Este conjunto de mensagens possui um enorme potencial em termos de visualização, visto descrever os elementos fundamentais da *performance* musical.

- As mensagens de *Control Change* e de *AfterTouch* permitem o envio de parâmetros de interpretação como o accionar de um pedal de expressão ou de alterações tímbricas como o ajuste da frequência de corte de um filtro, podendo por isso ajudar a conferir uma maior expressividade à interpretação musical, e como consequência, ao sistema visual.
- As mensagens de *Program Change* podem ter neste contexto de visualização um papel extremamente importante pois, ao ser escolhido um determinado *patch* no sistema sonoro, poderão ser accionados no sistema visual um conjunto de definições pré-programadas, armazenadas num *preset*. Torna-se assim possível desenvolver *presets* visuais que se aproximem de alguma forma da estética sonora de cada um dos *patches* dos instrumentos musicais.
- Outra mensagem MIDI que poderá ser útil é a de *MIDI Clock*, pois transmite a pulsação musical e, por consequência, dada a correlação pretendida, também a visual. Este tipo de mensagem poderá ser utilizado para efectuar o sincronismo de efeitos visuais periódicos com a música, colaborando para a sua cadência. É normalmente gerada por sequenciadores ou caixas de ritmo.
- Neste contexto, também as mensagens de *MIDI Time Code* (MTC) poderão ser úteis mas para sequenciação, de forma a sincronizar um eventual sequenciador sonoro com um sequenciador visual, responsável por gravar, editar e reproduzir toda a programação dos parâmetros de controlo manual, ou seja, não derivados da informação MIDI dos instrumentos musicais.

3.4.3 Esquemas de ligações

São várias as possibilidades de ligação do sistema sugerido, conforme indica a Figura 31. Na impossibilidade de utilização de um *interface* com várias portas, poder-se-á utilizar a funcionalidade de canais MIDI, configurando-se cada um dos instrumentos para comunicar em canais distintos. Neste caso, as ligações MIDI são efectuadas em cadeia entre os vários equipamentos utilizando as funcionalidades permitidas pela saída MIDI Thru. É de realçar no entanto, que o MIDI Thru apenas reenvia a informação recebida no conector MIDI In, não enviando por isso a informação gerada pela própria máquina. Para resolver situações como esta, em que necessitamos de acumular a informação recebida com a informação do próprio equipamento, muitos instrumentos possuem uma opção implementada em *software* que permite a comutação da função MIDI Thru para MIDI Merge. Na ausência desta possibilidade, será ainda possível recorrer a equipamentos como os MIDI Patch-Bays/MIDI Mergers [49].

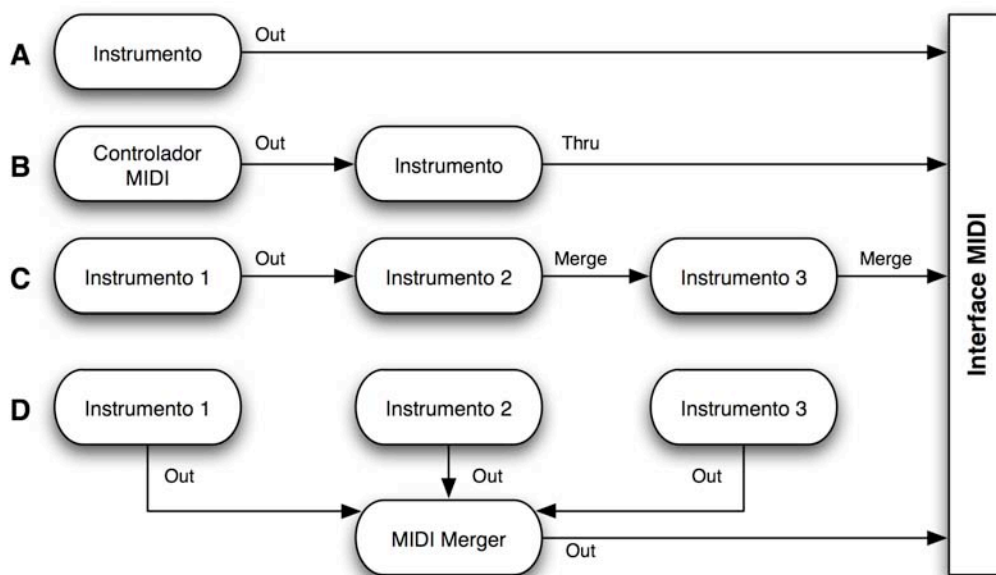


Figura 31 – Diferentes abordagens de envio de informação MIDI para o Sistema Visual

Uma vez que os cabos MIDI apenas suportam comprimentos até 15 metros, pode ser necessária uma alternativa para a transmissão desta informação entre o palco, onde os músicos se encontram, e o sistema de projecção vídeo, local onde preferencialmente se localizará o computador responsável pela construção da visualização. Uma solução possível será a utilização de um computador junto do palco que esteja equipado com um *interface* MIDI. Este deverá possuir o número de portas apropriado às necessidades do sistema sonoro¹ para que o computador possa receber a informação MIDI dos vários instrumentos e enviá-la através de uma tecnologia de rede disponível (Ethernet, Wi-Fi, entre outras) para o computador principal, localizado onde reside o sistema visual,² conforme descrito na situação B da Figura 32.

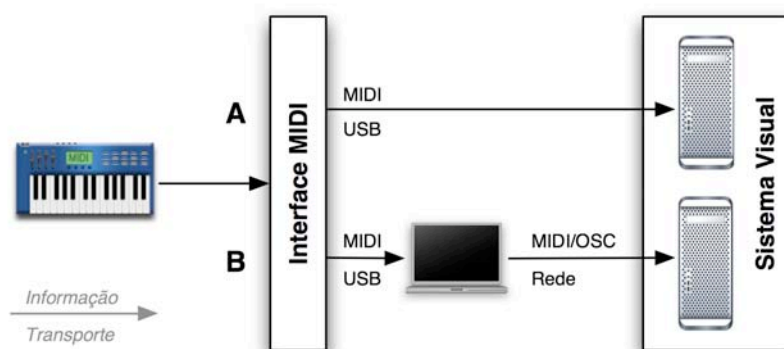


Figura 32 – Possibilidades de comunicação entre a informação MIDI e o Sistema Visual

¹ Existem inúmeras soluções de *interfaces* MIDI de marcas como a MOTU ou M-Audio, que fornecem configurações de uma a oito portas, utilizando actualmente o protocolo USB como forma de interligação com o computador.

² Para mais informações acerca da possibilidade de transmitir informação MIDI através de protocolos de rede, consultar o capítulo 3.2.1, “Transmissão de MIDI e Áudio entre aplicações locais e em rede,” pág. 80.

3.4.4 Programação

De forma a exemplificar a utilização de informação MIDI neste contexto de visualização sonora, decidimos incluir no sistema proposto um sequenciador de padrões rítmicos associado a um popular módulo de sons em *hardware*, o Roland JV-1080. Tirámos partido do sequenciador de padrões rítmicos integrado na aplicação Digital Performer 6 (Figura 33) e enviámos as mensagens MIDI de nota para o módulo de sons e, através da função *MIDI Thru* do mesmo, também para o computador responsável pela síntese visual, onde se encontra instalado o Max/MSP/Jitter. Dada a sua origem, este ambiente de programação possui um conjunto de objectos bastante completo no que diz respeito à recepção, tratamento e envio de informação MIDI, pelo que não tivemos qualquer dificuldade no processo de programação do sistema proposto. Dos objectos utilizados para receber as várias informações MIDI podemos destacar o [notein], o [pgmin], o [ctlin] e o [rtin].

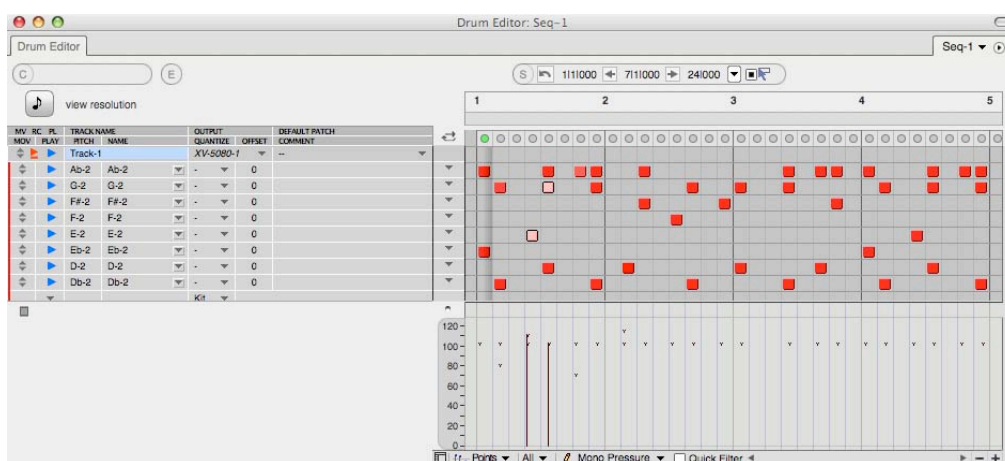


Figura 33 – Sequenciador de padrões rítmicos da aplicação Digital Performer 6

A descrição de altura de cada nota tocada define qual o elemento do *kit* de bateria/percussão accionado (bombo, tarola, pratos, etc). Associado ao valor da nota está também a caracterização da dinâmica do ataque. O sistema visual deverá estar já pré-configurado para representar cada uma das peças da bateria (notas) através de elementos visuais distintos, quer pelo posicionamento no espaço visual, quer pela aparência (cor, forma, textura, etc).

Foi ainda considerada a possibilidade de automatizar a mudança de programas (*presets*) sonoros e visuais através da inclusão de mensagens de *Program Change*. Assim, quando alterarmos a selecção do *kit* de bateria do módulo de sons, poderemos obter automaticamente uma alteração da componente visual através de *presets* previamente programados. Incorporamos ainda a possibilidade de ajuste do volume do módulo de sons no canal MIDI utilizado através de mensagens de *Control Change 7*. O envio deste parâmetro para o sistema visual poderá, por exemplo, definir o grau de opacidade dos objectos representados. De forma a obtermos no sistema visual a pulsação musical (*andamento*), poderíamos servir-nos da mensagem de *MIDI Clock*, presente na maioria dos sequenciadores MIDI. É de realçar que mesmo em contextos de música ao vivo, particularmente em contextos de música electrónica, *pop* ou *rock*, é cada vez mais frequente o recurso a

metrónomos a partir de equipamentos electrónicos como base temporal para a actuação musical, pelo que este tipo de informação MIDI se encontra muitas vezes disponível.

Situações como a que propomos atrás são bastante adequadas à tecnologia MIDI uma vez que não existe uma tentativa de representação das características tímbricas, mas apenas de um simples accionamento e caracterização de notas ou outros parâmetros relativamente simples. Em implementações mais complexas baseadas em MIDI, poderá tornar-se necessário contemplar situações como o eventual uso do pedal de sustém ou até do controlador de *pitchbend*. Se no primeiro caso é relativamente fácil atrasar o accionamento das mensagens de *note off* até à libertação do pedal, já neste último, torna-se impossível descrever a quantidade de desvio da afinação uma vez que esta informação se encontra programada dentro da estrutura do próprio sintetizador e varia normalmente de programa para programa. Uma grande parte do inconveniente do MIDI prende-se exactamente com este facto: a impossibilidade de acesso à estrutura interna do sintetizador. Um outro exemplo prático com reflexo na abordagem por nós desenvolvida é o facto do par de comandos *note on/note off* não ser de todo suficiente para descrever a amplitude de uma nota ao longo do tempo; o som de um instrumento percussivo, com um rápido decaimento, poderá extinguir-se muito antes do comando *note off* ser executado (ou o oposto), prejudicando assim a relação audiovisual pretendida. Uma solução parcial para este problema particular foi por nós testada ao permitir no sistema visual a definição de um valor de decaimento para cada um dos objectos sonoros a representar, de forma a permitir uma maior adequação com o tempo de decaimento dos respectivos sons.

Dado este tipo de limitações constatadas e a viabilidade da alternativa a seguir apresentada, baseada no acesso aos parâmetros da estrutura interna do sintetizador sonoro como base para o mapeamento audiovisual, decidimos não explorar de forma exaustiva a implementação MIDI. Consideramos, no entanto, que para determinado tipo de situações, pela sua descrição da execução musical, poderá configurar-se como a solução mais interessante. Por fim, é importante referir que poderão existir soluções práticas capazes de minimizar algumas das limitações inerentes a esta tecnologia, nomeadamente, pela utilização conjunta deste protocolo com o próprio resultado sonoro (áudio digital) de modo a obter uma mais completa caracterização musical e tímbrica.

3.5 Visualização de Síntese Sonora

O sistema sonoro concebido na aplicação Reaktor é constituído por um conjunto de equipamentos virtuais, cada um com determinada função em termos sonoros, mas também com repercussões ao nível visual, dado que este serve simultaneamente como principal fonte de informação para o sistema visual. Como fonte sonora principal construímos um comum sintetizador baseado em síntese substractiva. Foi incluída ainda uma fonte sonora adicional (*sample player*) como mero suporte para a ilustração das possibilidades colaborativas no âmbito da visualização. Estes encontram-se ligados a uma mesa de mistura virtual onde se pode efectuar o ajuste de níveis, panorâmicas, e envio para canais auxiliares. Estão ainda ligados à mesa de mistura dois processadores de efeitos através da comum técnica de *send/return*¹: um processador de *delay* (Dual Sync Delay) e um de reverberação (Space Master 2). De todos estes equipamentos virtuais, o sintetizador foi aquele no qual concentrámos maior atenção. A mesa de mistura serve não só para definir o equilíbrio dos vários intervenientes musicais, como também das suas respectivas representações visuais. As alterações sonoras fruto dos processadores de efeitos aqui presentes encontram-se também representadas visualmente. De seguida, descreveremos de forma resumida algumas particularidades referentes a cada um dos componentes do sistema sonoro concebido.

3.5.1 Componentes do Sistema Sonoro

Sintetizador

A principal fonte sonora desenvolvida, o sintetizador, baseia-se num tradicional modelo de síntese substractiva. Ao analisarmos a Figura 34 podemos ver de forma esquemática a composição do sintetizador construído. Logo à entrada, a informação passa por um arpegiador e uma série de opções relativas ao modo de funcionamento (monofónico, polifónico, etc.). O oscilador permite seleccionar um conjunto de formas de onda (quadrada, triangular, dente de serra, sinusoidal...) e serve de base para toda a construção sonora. Quer a amplitude das notas, quer a frequência de corte do filtro (LPF) podem ser moduladas por envolventes sonoras e LFOs dedicados. Posteriormente, o resultado áudio é enviado para um conjunto de processadores de efeitos, de forma a enriquecer o som inicial. O efeito *chorus* transforma o sinal em *stereo* e a partir daí o processamento efectuado mantém-se nesse modo. Existe ainda a possibilidade de, já fora do sintetizador, através da mesa de mistura acrescentar os tradicionais efeitos de *reverb* e de *delay*.

¹ Técnica onde se efectua o envio do áudio por meio dos auxiliares, e o retorno para canais específicos.

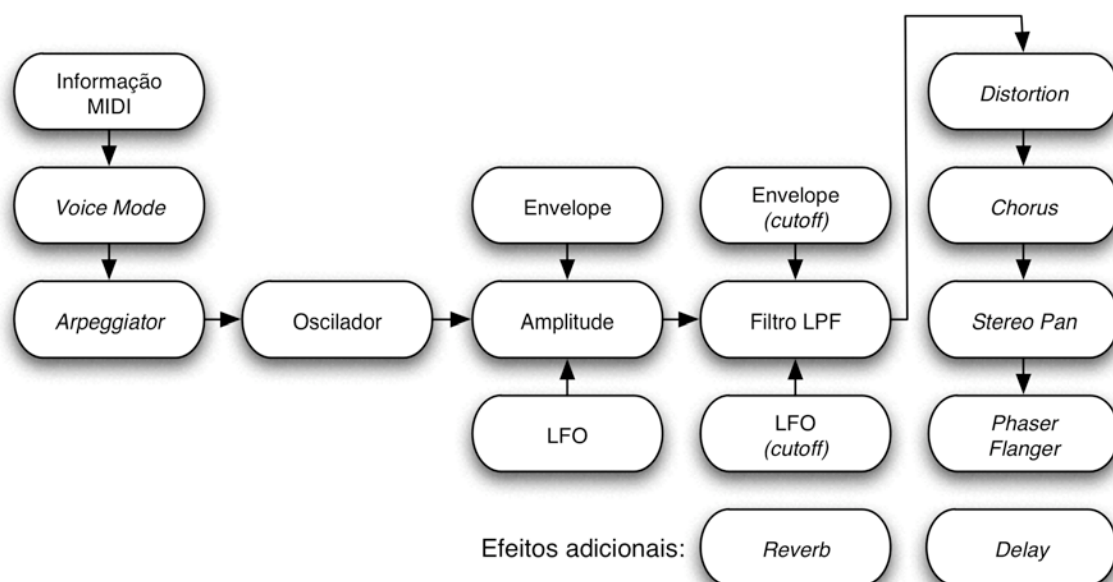


Figura 34 – Estrutura do Sintetizador

Mesa de Mistura

A mesa de mistura por nós utilizada (Figura 35) possui 4 canais estereofônicos, cada um com possibilidades de envio para 2 canais auxiliares (*Send 1* e *Send 2*), definição de panorâmica, função *Solo* e *On/Off*, e ajuste de nível através de um *fader*. O barramento de envio *Send 1* está configurado como *post fader*, enquanto que o do *Send 2* como *pre fader*. O primeiro encontra-se ligado a um processador de *delay*, o “Dual Sync Delay” e o segundo a um processador de reverberação, o “Space Master 2”. Na secção de *Master*, para além do óbvio *Master Fader*, encontram-se ainda presentes 2 canais estereofônicos para o retorno de efeitos (ER1, ER2) com ajuste de nível e função *Solo*.

A mesa de mistura possui aqui um papel extremamente importante na medida em que é responsável pelo equilíbrio de todo o sistema, ou seja, do sistema sonoro, mas também do visual. Por exemplo, ao efectuarmos *mute* ao canal do sintetizador, deixaríamos de o ouvir mas também de o visualizar. No caso das alterações de volume dos canais (*faders*), as mudanças não serão tão drásticas, uma vez que o nível de transparência das respectivas representações visuais é progressivamente ajustado. Também os auxiliares de envio para os processadores de efeitos e respectivos retornos se encontram relacionados com o sistema visual. No caso do efeito de reverberação, o parâmetro responsável pela definição do tempo de reverberação encontra-se relacionado com a quantidade de *feedback* do sistema visual, proporcionando, tal como no campo auditivo, uma persistência no tempo. No caso do efeito de *delay*, este possui uma representação visual similar, ou seja, mediante a utilização de um *buffer* com determinado tamanho, poderá ser adicionado ao sistema visual um eco, que poderá ou não ser re-alimentado.

Dual Sync Delay

O efeito de atraso do som, mais conhecido pelo termo “*delay*”, é um efeito extremamente simples, mas a sua utilização pode potenciar consideravelmente o som de um sintetizador. Tem por base o fenómeno físico do “eco”, repetição do som original com um determinado atraso causado pela reflexão sonora em alguma superfície. No entanto, os processadores de efeitos não se limitam a efectuar apenas este atraso; é habitual a possibilidade de definir muitas outras opções como a filtragem, a re-alimentação ou o número de ecos independentes e o seu posicionamento no espaço estereofónico. Este efeito serve de base para outros efeitos como a reverberação, o *chorus* ou o *flanger*. O processador *Dual Sync Delay* (Figura 36), para além das excelentes capacidades em termos de alteração tímbrica, permite a especificação do atraso através de valores de tempo absoluto ou de tempo musical (informação proveniente dos comandos *MIDI Clock*). Esta última opção permite a sincronização do atraso com, por exemplo, o arpegiador ou com os LFOs presentes no sintetizador. No caso concreto do nosso sistema sonoro, em vez de ser parte constituinte do sintetizador, este integra-se com a mesa de mistura através da técnica de *Send/Return*. Assim sendo, o parâmetro *Mix*¹ deste processador deverá estar configurado com o valor 1 (som 100% processado), visto que o som original já se encontra presente na mesa de mistura. Estando o auxiliar 1 (*Send 1*) configurado como *post fader*, o nível de envio para este processador é sempre proporcional ao nível de envio para o *Master*.

Space Master 2

A reverberação é um fenómeno sonoro caracterizado pela permanência do som num espaço por determinado período de tempo, e por isso é responsável pelo reconhecimento auditivo desse mesmo espaço. Esta é composta pelas múltiplas reflexões sonoras que chegam ao receptor com diferentes atrasos relativamente ao som directo (som original), consoante a extensão do percurso percorrido. A sonoridade geral de uma reverberação é dada pela configuração dos seus parâmetros, dos quais destacamos as relações de intensidade, atraso com o som directo, as primeiras reflexões e o som reverberante. Estes parâmetros afectam o som do ponto de vista da localização da fonte sonora, da clareza, intimidade e difusão, alterando radicalmente a nossa percepção do espaço envolvente.

Dada a importância deste fenómeno acústico, praticamente qualquer sintetizador construído na última década possui processadores de efeitos integrados que disponibilizam algoritmos de reverberação. Podemos afirmar que este aspecto se tornou um elemento importante na sonoridade geral do sintetizador. Esta é a razão da inclusão do processador *Space Master 2* (Figura 37) como complemento ao sintetizador sonoro por nós criado. Dado o conceito de “estúdio virtual” proporcionado pelo Reaktor, optámos por não o integrar directamente no sintetizador, como acontece em qualquer implementação de *hardware*, mas sim simular uma situação tradicional de estúdio, onde o processador de reverberação se liga pela técnica de *Send/Return* à mesa de mistura, possibilitando que cada instrumento a ela ligado possa beneficiar do mesmo. Nesta técnica,

¹ Neste contexto, *Mix* significa a relação entre o som original e o som processado.

o parâmetro *dry/wet* deve estar configurado para “100% wet”, ou seja, a saída do processador não deverá conter uma mistura do som original, dado que este já se encontra presente num canal da mesa de mistura.

Uma abordagem possível de relacionamento visual deste efeito poderá ser obtido mediante a utilização de um sistema de *feedback* dos gráficos gerados, de forma a simular a persistência temporal do efeito sonoro. Tentando simular o fenómeno de difusão, poderá ser contemplado ainda o uso de um efeito de *blur*, desfocando progressivamente os gráficos gerados.



Figura 35 – A mesa de mistura



Figura 36 – O processador *Dual Sync Delay*

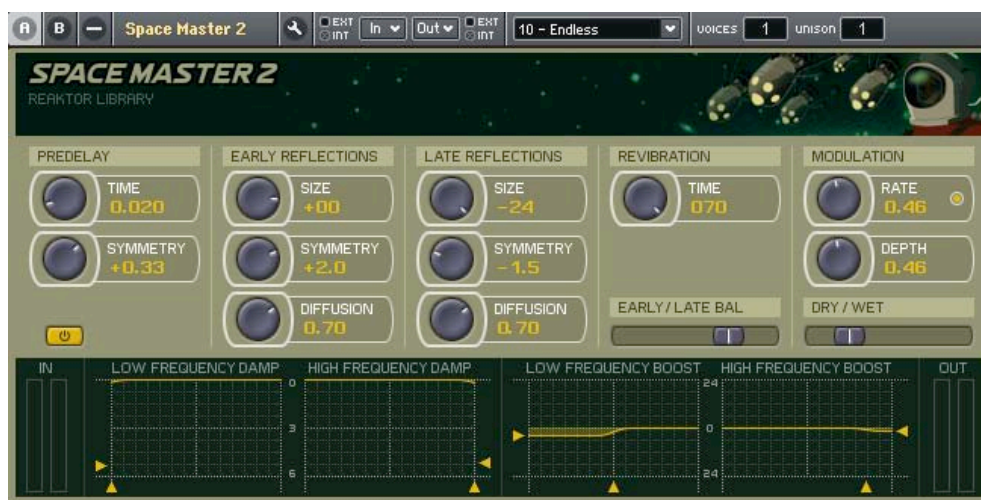


Figura 37 – O processador *Space Master 2*

3.5.2 Comunicação entre os sistemas através de OSC

A comunicação entre o sintetizador sonoro e o sistema visual possui, por motivos óbvios, uma importância fulcral no contexto deste trabalho. Esta comunicação é assegurada pelo protocolo OSC,¹ através de qualquer tecnologia de rede que suporte a arquitectura protocolar TCP/IP. No nosso caso concreto, existe a necessidade de compatibilizar a comunicação entre as duas aplicações escolhidas para a implementação do sistema proposto: o Reaktor para a síntese sonora, e o Max/MSP/Jitter para a síntese visual. No Reaktor é necessário definir o endereço IP de destino da informação e o respectivo porto utilizado. No Max/MSP/Jitter, receptor da informação, é necessário especificar apenas o porto utilizado. Desta forma, torna-se possível a comunicação entre o sistema sonoro e visual, desde que estes estejam em sintonia ao nível dos parâmetros utilizados. Nesta secção pretendemos referir as dificuldades encontradas e as metodologias utilizadas para efectuar a comunicação entre estas duas aplicações e enunciar o mapeamento de parâmetros por nós efectuado.

Sintetizador Sonoro (Reaktor)

No Reaktor, existem duas formas de enviar parâmetros pelo protocolo OSC:

1. a ligação do parâmetro a enviar ao módulo [OSC Send];
2. a activação da configuração de envio, presente nos vários módulos que possuam uma representação visual no painel (*faders*, *knobs*, botões, listas, etc.)

Destes dois métodos, este último possui a vantagem de não ter que ser incluído propositadamente para o efeito. Detectámos contudo algumas desvantagens que nos levaram a utilizar preferencialmente o primeiro método. A primeira prende-se com o facto do nome dos elementos do painel ser automaticamente utilizado como nome do parâmetro OSC enviado. Utilizando este método, cada elemento do *interface* deverá possuir um nome único, facto que em algumas situações se poderá tornar inconveniente em termos ergonómicos; por vezes vários elementos do *interface*, em partes distintas do sistema sonoro, sugerem a utilização do mesmo nome. A segunda desvantagem constatada é a de todos os elementos do painel enviarem parâmetros OSC formatados como *floats* com valores compreendidos entre 0 e 1; mesmo que os valores manipulados pelo controlador em causa possuam outro registo (ex: -120 a 50), o Reaktor escala estes valores.² Se em alguns casos de correlação com o sistema visual este procedimento pode até ser uma vantagem, noutros casos vem dificultar o processo de compatibilização.

¹ Ver “O OpenSound Control,” pág. 49

² Esta limitação tem muito provavelmente como justificação a possibilidade do Reaktor funcionar em modo de *plug-in* VST, AU, etc. em contextos de sequenciação. A especificação VST requer que todos os parâmetros de controlo do *plug-in* possuam valores *float* compreendidos entre 0 e 1.

A maioria dos módulos do Reaktor pode funcionar em dois modos distintos: o modo monofónico e o modo polifónico.¹ Este facto facilita imenso a programação de sintetizadores polifónicos, pois esta é gerida automaticamente pelo programa. Infelizmente, na actual versão do Reaktor não é possível utilizar o objecto [OSC Send] no modo polifónico. Dado que alguns parâmetros mapeados são transmitidas utilizando este modo (o sintetizador programado é polifónico), necessitámos de proceder previamente à divisão da informação relativa a cada uma das vozes, de forma a enviá-las como parâmetros OSC distintos. Os parâmetros que foram sujeitos a esta alteração foram a indicação das notas tocadas, da sua amplitude (incluindo já o envelope sonoro), e da frequência de corte do filtro. A razão pela qual enviamos esta última de forma separada (por voz) e não de forma global prende-se com o facto desta depender, para além do controlo dedicado (*cutoff*), da nota tocada e/ou da sua intensidade (amplitude).

Existem ainda outros parâmetros como a variação cíclica dos LFOs, a quantidade de cada um dos efeitos, dos níveis da mesa de mistura ou a informação de pulsação musical (*MIDI Clock*), que foram também enviados por OSC tendo em vista a sua possível utilização por parte do sistema visual (para uma informação mais detalhada consultar a Tabela 8, pág. 108).

Finalmente, é importante referir aqui um outro problema constatado na compatibilização do sistema sonoro com o sistema visual. O Reaktor possui internamente três tipos de informação em termos de cadência temporal: eventos pontuais, eventos com uma frequência de amostragem de controlo (*control rate*) e eventos com uma frequência de amostragem de áudio (*sample rate*). Esta última é utilizada apenas para o áudio. Alguns parâmetros como o accionar de uma função serão, obviamente pontuais. Os eventos que utilizam a frequência de controlo poderão ser, por exemplo, a descrição de um envelope sonoro, ou a descrição da variação de um LFO. De forma a obter os melhores resultados sonoros, essa frequência (que pode ser configurável entre os 25 e os 3200 Hz) deverá ser relativamente elevada.

O débito de envio dos parâmetros OSC é coincidente com esta frequência de controlo seleccionada. Embora esta forma de funcionamento seja lógica, no contexto em causa, pode tornar-se prejudicial ao funcionamento do sistema proposto pois, quando a frequência de controlo, por uma questão de qualidade sonora, for consideravelmente elevada (ex: 3200 Hz), poderão acontecer dois problemas distintos na comunicação com o sintetizador visual:

1. o débito de informação a ser enviada por rede poderá revelar-se excessivo, dependendo da quantidade de parâmetros a ser enviados e do tipo de tecnologia de rede utilizada.
2. o desempenho do sintetizador visual poderá ser prejudicado, uma vez que recebe e processa as informações de controlo com uma frequência desnecessária em termos de percepção visual. Devido

¹ Na visualização da estrutura (modo de construção do Reaktor) a luz amarela representa o modo polifónico, enquanto que a luz laranja representa o modo monofónico

à frequência máxima com que gera as imagens e à própria percepção visual, este não necessitaria seguramente de receber informação com uma frequência acima dos 100 Hz.

A solução para o primeiro problema levantado é o de utilizar tecnologias de rede com um bom desempenho em termos de débito binário (ex: *Gigabit Ethernet*). Por outro lado, mesmo tendo como preocupação a qualidade sonora, não existe uma grande necessidade de utilizar frequências de controlo tão elevadas. A solução para o segundo problema será o de implementar um sistema no sintetizador visual em que a frequência de processamento não seja coincidente com a frequência de recepção das informações de controlo. De forma a otimizar o envio de informação através da rede, decidimos ainda efectuar uma redução dos dados de controlo enviados. Essa redução assentou na filtragem de toda a informação redundante. Para isso, recorreremos à inclusão do módulo [Step Filter] antes do envio dos dados por OSC (excepto os dados de controlo pontuais); este apenas deixa passar informação quando esta difere da anterior, diminuindo assim consideravelmente a quantidade de dados transmitidos.

Sintetizador Visual (Max/MSP/Jitter)

Sendo o Max/MSP/Jitter um verdadeiro ambiente de programação, não possui algumas das restrições detectadas na aplicação Reaktor, atrás mencionada. Permite por isso uma maior liberdade na utilização da informação de controlo proveniente do sintetizador sonoro através do protocolo OSC [65]. Este facto revela-se extremamente importante na medida em que nos permite melhor compatibilizar os dois sistemas. Por outro lado, enquanto que o suporte para o protocolo OSC encontra-se implementado de raiz no Reaktor, para podermos utilizar o mesmo no ambiente Max/MSP/Jitter, necessitámos de recorrer a objectos externos (*externals*). Esta funcionalidade não se encontra ainda presente na actual versão do referido *software*. O desenvolvimento destes *externals* para MacOS X é da responsabilidade de Matt Wright e o *port* destes para Windows é da autoria de Rémy Muller e de Emmanuel Flety.

Estes são os objectos incluídos para a utilização do protocolo OSC no Max/MSP:

[udp-read] e [udp-write]	Estes dois objectos permitem enviar e receber dados por ethernet utilizando o protocolo UDP (através da API “Winsock” do Windows).
[opensoundcontrol]	Objecto responsável pela conversão de dados Max/MSP em pacotes OSC e vice-versa.
[osc-route]	Objecto responsável pelo routing dos vários tipos de mensagens provenientes do OSC para destinos diferenciados.

Durante a execução desta tese, a Cycling’74 lançou dois novos objectos, [udpreceive] e [udpsend], substituindo as funcionalidades dos objectos [udp-read] e [udp-write] acima mencionados, e ainda algumas funcionalidades incluídas no [opensoundcontrol]. Depois de testados exaustivamente os vários objectos referidos, encontrámos alguns problemas na recepção dos parâmetros do Reaktor no ambiente Max/MSP.

Esses problemas estavam relacionados com a versão utilizada do objecto [opensoundcontrol]¹. No entanto, conseguimos resolver a situação utilizando apenas o [udpreceive], pois como foi referido anteriormente, este objecto possui já a capacidade de interpretar mensagens OSC, embora com algumas limitações que para o nosso caso particular não se revelaram como significativas. O outro objecto utilizado conjuntamente com o [udpreceive] foi o [osc-route], de forma a obter os vários parâmetros enviados por OSC.

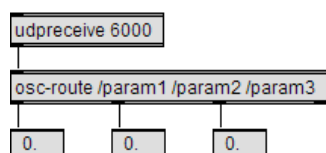


Figura 38 – Método de recepção dos dados OSC

3.5.3 Relações estabelecidas entre o Sintetizador Sonoro e o Visual

Uma vez que no Reaktor é-nos possível aceder a todos os elementos que contribuem para a construção sonora, desde a informação MIDI inicial até à saída do áudio gerado, torna-se extremamente importante perceber que tipo de parâmetros será conveniente utilizar de forma a relacionar o sistema sonoro com o visual. Enquanto que o mapeamento a um nível demasiado superficial (áudio gerado) não possibilitaria uma suficiente descrição de parâmetros, também o mapeamento a um nível demasiado complexo apresenta os seus problemas. Este obrigar-nos-ia a implementar no sintetizador visual todos os elementos constituintes da estrutura utilizada pelo sintetizador sonoro. Como exemplo, se cada um dos controladores ADSR responsáveis pela definição de uma envolvente sonora fossem utilizados directamente neste mapeamento, tornar-se-ia necessário que, para além do Reaktor, também o Max/MSP/Jitter tivesse que recriar esta envolvente sonora resultante. Analisando de forma global o sistema, encontramos nesta situação uma clara redundância ao nível da funcionalidade, tendo como consequência uma complexidade acrescida e uma redução na capacidade de processamento, sem que daí advenha qualquer vantagem. Em nosso entender, a abordagem correcta será, neste exemplo, a de fornecer ao sistema visual a informação obtida à saída da envolvente sonora. Generalizando, a escolha relativa ao tipo de parâmetros a utilizar tem de ser ponderada caso a caso, mediante o grau de independência de controlo que necessitamos de obter para a correlação com o sistema visual. Podemos observar na Figura 39 que a descrição da amplitude e da afinação das notas tocadas passa por uma série de processos que, apesar de não se encontrarem relacionados directamente com o sistema visual, acabam por fazê-lo de forma indirecta, uma vez que influenciam os parâmetros enviados por OSC.

A Tabela 8 a seguir apresentada descreve o relacionamento efectuado entre parâmetros do Sintetizador Sonoro e parâmetros do Sintetizador Visual. A selecção dos parâmetros utilizados obedeceu ao critério acima

¹ A versão utilizada deste pacote foi a 1.5.1.

descrito. O tipo de informação comunicada é em alguns casos pontual (ex: níveis de efeitos, ressonância do filtro), e noutros contínua (definição da envolvente sonora, LFOs).

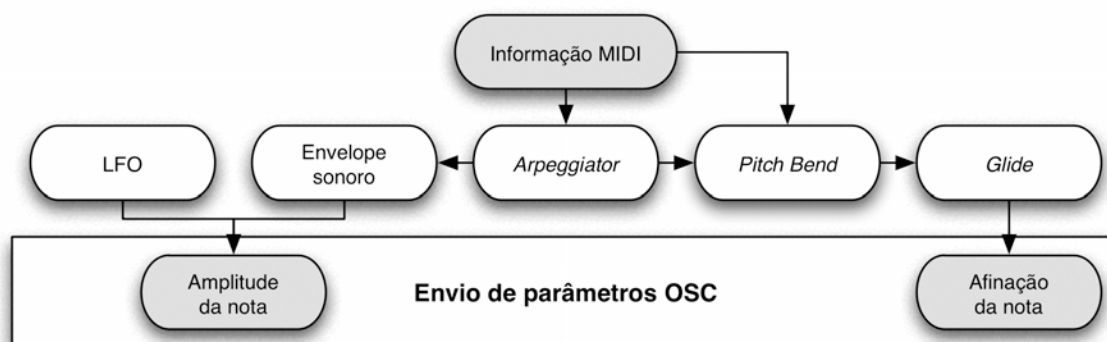


Figura 39 – Opções de envio dos parâmetros das notas (amplitude e afinação)

Tabela 8 - Principais mapeamentos efectuados por OSC entre parâmetros sonoros e visuais

Contr.	Parâmetro sonoro	Parâmetro (OSC)	Parâmetro visual
Teclas	Velocidade de ataque das notas	attack	Accionamento e quantificação do efeito de reposicionamento
Teclas	Afinação da voz	p1 – p8	Frequência (onda) Cor (tonalidade)
Teclas	Amplitude da voz	a1 – a8	Amplitude (onda) Transparência (onda)
Pitch Bend	Alteração da afinação	(p1 – p8)	Frequência (onda) Cor (tonalidade)
Mod. Wheel	Amplitude do LFO relativo ao Filtro	lfof	Amplitude do LFO relativa à posição no eixo Y da âncora de rotação/escala
R1	Frequência de Corte do Filtro	f1 – f8 Cutoff	Luminosidade (onda) Desfocagem (<i>shader gaussian.2p</i>)
R2	Ressonância do Filtro	Reson	Saturação (onda) Rotação
R3	Amplitude do LFO relativo à Amplitude	lfoa	Amplitude do LFO relativa à posição no eixo X da âncora de rotação/escala
R4	Amplitude do LFO relativo ao Filtro	lfof	Amplitude do LFO relativa à posição no eixo Y da âncora de rotação/escala
R5-R8	ADSR da Amplitude	(a1 – a8)	ADSR da Amplitude (onda)
S1	Frequência do LFO1	(a1 – a8)	Frequência do LFO relativa à posição no eixo X da âncora de rotação/escala
S2	Frequência do LFO2	(f1 – f8)	Frequência do LFO relativa à posição no eixo Y da âncora de rotação/escala
S3	Frequência do LFO relativo ao Stereo	stereolfo	Frequência do LFO relativa à rotação no eixo Y
S4	Amplitude do LFO relativo ao Stereo	stereolfoamp	Amplitude do LFO relativa à rotação no eixo Y
S5	Nível de Distorção	distort	Nível de saturação e contraste da cor (<i>shader sharpen</i>)
S6	Nível de Chorus	chorus	Desfazamento horizontal dos 2 planos OpenGL
S7	Nível de Flanger	ampflanger	Nível de Distorção (<i>shader sinefold</i>)
S8	Frequência do Flanger	lfoflanger1	Frequência da distorção (<i>shader sinefold</i>)

3.6 O Sistema Visual

No desenvolvimento deste sintetizador responsável por dar forma visual ao som gerado, não pretendemos ser particularmente inovadores em termos de estética final, uma vez que partimos de algumas técnicas comuns no domínio da visualização de áudio. Aliás, dado não possuímos qualquer experiência prévia neste domínio, seria difícil outro tipo de abordagem. A nossa preocupação foi apenas a de construir um modelo visual relativamente simples, que nos permitisse explorar correlações entre os sentidos auditivo e visual. Desta forma, o resultado final do ponto de vista artístico não será o essencial, embora sejamos da opinião que, se bem utilizado, este modelo permite gerar ambientes visuais interessantes e bastante diferenciados.

De seguida, depois de formulado o conceito geral do sistema visual proposto, descrevemos de uma forma mais pormenorizada os principais aspectos relativamente à implementação da visualização das três fontes de que dispomos: o Áudio Digital, o MIDI e a Síntese Sonora.

3.6.1 Conceito

David Gibson, no seu livro intitulado *The Art of Mixing* [66] tendo como objectivo demonstrar uma abordagem visual ao processo de mistura de áudio (Figura 40) sugere a representação visual do volume sonoro de determinado instrumento pela profundidade ou distância (eixo Z) do objecto que o representa. Para além da mais óbvia relação entre o posicionamento estereofónico de um som (panorâmica) e a localização visual no eixo horizontal (X), Gibson relaciona ainda as frequências sonoras com o eixo vertical (Y).



Figura 40 – Representação visual da mistura de um tema musical de estilo *Blues*

A ideia subjacente ao sistema visual por nós proposto é a de um “palco tridimensional” onde cada instrumento sonoro seja representado visualmente por um ou mais objectos, cada um contribuindo de formas distintas para o resultado visual, tal como acontece no campo auditivo, onde cada instrumento possui uma função musical específica. Assim como o som de cada um dos instrumentos se funde no espaço acústico,

também o resultado visual será uma combinação destes vários elementos, sendo a representação visual de cada instrumento resultado da sua sonoridade ou função musical. Esta abordagem vai de encontro à opinião de Tom DeWitt, que considera a imagem total como uma espécie de orquestra visual [61].

Existe uma diferença fundamental entre a aparência acústica e a visual: os objectos sonoros podem ser considerados transparentes, na medida em que um som não impede outro de ser ouvido.¹ Podemos perceber todos os sons em simultâneo, ao contrário do que acontece no mundo visível, onde um objecto opaco não permite a propagação da luz, e com isso a nossa percepção de outros objectos posicionados atrás de si. Assim, podemos considerar que os objectos visuais, dependendo do seu posicionamento no espaço relativamente ao observador, “mascaram-se” mutuamente, impedindo uma visão completa do universo visual. Dado o intuito de efectuar uma representação visual dos objectos sonoros, torna-se necessário ultrapassar esta diferença fundamental entre os dois domínios, ao fazer com que a representação visual partilhe esta característica sonora. No decorrer desta tese, foram analisadas e implementadas várias abordagens para satisfazer esse propósito:

- atribuição de uma determinada percentagem de transparência aos objectos do ambiente visual, impedindo assim que estes ocultem totalmente os objectos posicionados atrás de si;
- inclusão de um *alpha channel* nas texturas *bitmap* que permita a definição das áreas de transparência/opacidade dos objectos onde são aplicadas;
- junção de todos os objectos numa imagem final através de vários métodos de composição (ex: *darken, lighten, multiply, screen, overlay*), como os presentes em *software* de edição de imagem.²

Na Figura 41 podemos observar de forma simplificada alguns elementos constituintes do sistema visual. A informação musical poderá ser representada visualmente por uma série de planos alinhados com o ângulo de visão ou com objectos tridimensionais. Nestes podem ser mapeadas texturas geradas através de várias técnicas de processamento de imagem. Como elemento gerador inicial, que podemos comparar em termos funcionais aos osciladores ou geradores de um sintetizador sonoro, poderemos utilizar uma forma de onda representativa do áudio tal como apresentada na Figura 42, por um osciloscópio (domínio temporal), ou a curva definida pelo espectro sonoro (domínio frequencial, obtido através de um processo de FFT). No caso da informação musical se basear em MIDI ou na síntese sonora, optamos por utilizar uma forma de onda básica como a sinusoidal. Aqui, cada nota tocada gera um destes elementos básicos. Ao contrário da visualização de áudio, não é aqui considerado o timbre sonoro, apenas a frequência fundamental das notas tocadas. Visto que toda a informação do sintetizador sonoro é potencialmente utilizável, os vários elementos responsáveis pela alteração sonora podem estar representados visualmente através de outros parâmetros

¹ exceptuamos aqui o fenómeno de “Efeito de Máscara” (Henrique, 2002) onde uma onda sonora pode ser “anulada” em termos de percepção quando ouvida em conjunto com outra de frequência próxima à primeira, mas de maior intensidade.

² Para mais informação acerca do processamento efectuado pelos vários modos de composição, consultar <http://www.northlite.net/ps/blend.htm>

como a cor, transparência, posição, rotação, ou qualquer outro processamento adicional. Em qualquer destas abordagens, a escolha da cor do gerador é um parâmetro essencial para o resultado final da visualização. Para uma maior sensação de correlação audiovisual, esta escolha poderá estar relacionada com aspectos musicais.

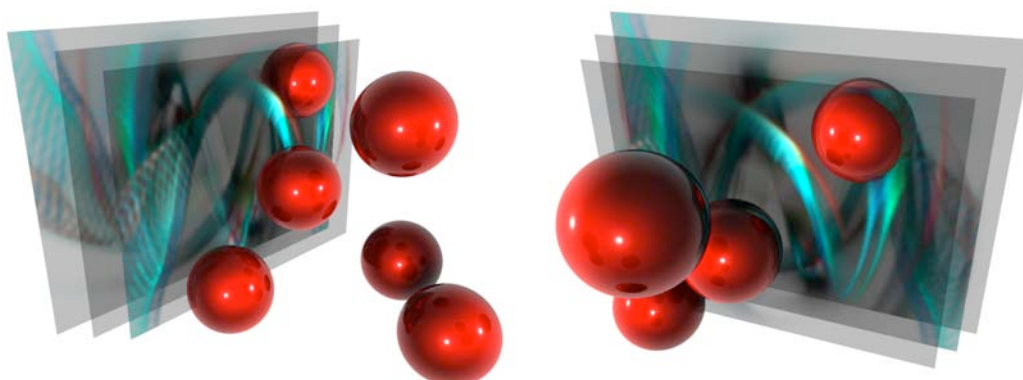


Figura 41 – Perspectivas do palco tridimensional do Sistema Visual



Figura 42 – Gerador inicial tipo osciloscópio (domínio temporal) com dois aspectos distintos: linha e área

Como acontece em qualquer sintetizador sonoro, ao *gerador* estão sempre associados módulos de processamento denominados de *modificadores*. Neste caso, convertemos os gráficos gerados (originalmente em OpenGL) para uma representação *bitmap* de forma a processá-la através do recurso a *shaders* programáveis (GPU) com vários efeitos, dos quais destacamos uma cadeia de *feedback* com processamento de rotação, escala e deformação (em grande parte responsável pelo aspecto final), desfocagem (*blur*), distorção, nitidez (*sharpen*) e saturação de cor. A textura resultante de todo este processamento é posteriormente mapeada em objectos bi ou tridimensionais (OpenGL), normalmente em planos alinhados com o ângulo de visão. Contudo, estes podem ser totalmente configuráveis em termos de posicionamento, escala e rotação. Visto que as texturas geradas incluem um canal de transparência, e que a forma de composição das mesmas poderá ser aditiva, torna-se possível visualizá-las em simultâneo (Figura 43).

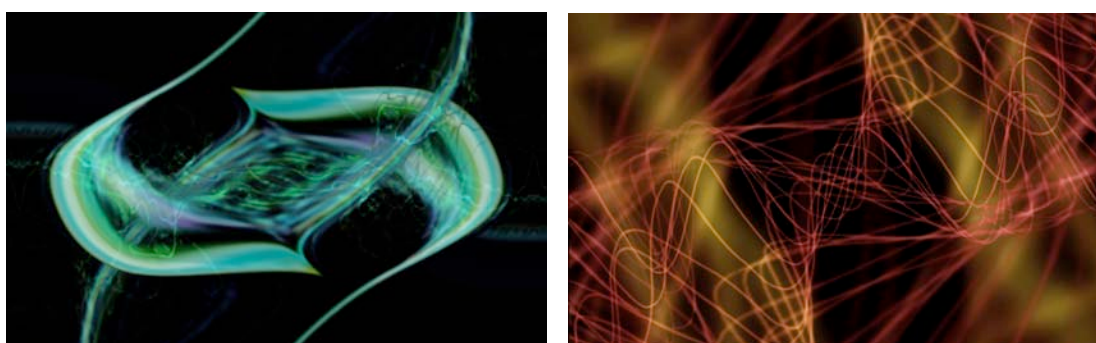


Figura 43 – Imagens compostas por dois planos distintos com recurso a *alpha channel* (esq.) e adição (dir.)

Na Figura 44 podemos observar uma visão esquemática simplificada de uma das primeiras implementações deste sistema visual, referindo o tipo de processamento envolvido (CPU/GPU) e o objecto do processamento: geometria, a cinzento, ou textura, a branco. Esta implementação mista de processamento entre o CPU e o GPU foi posteriormente objecto de optimização, tendo-se conseguido efectuar a totalidade do processamento pelo GPU.

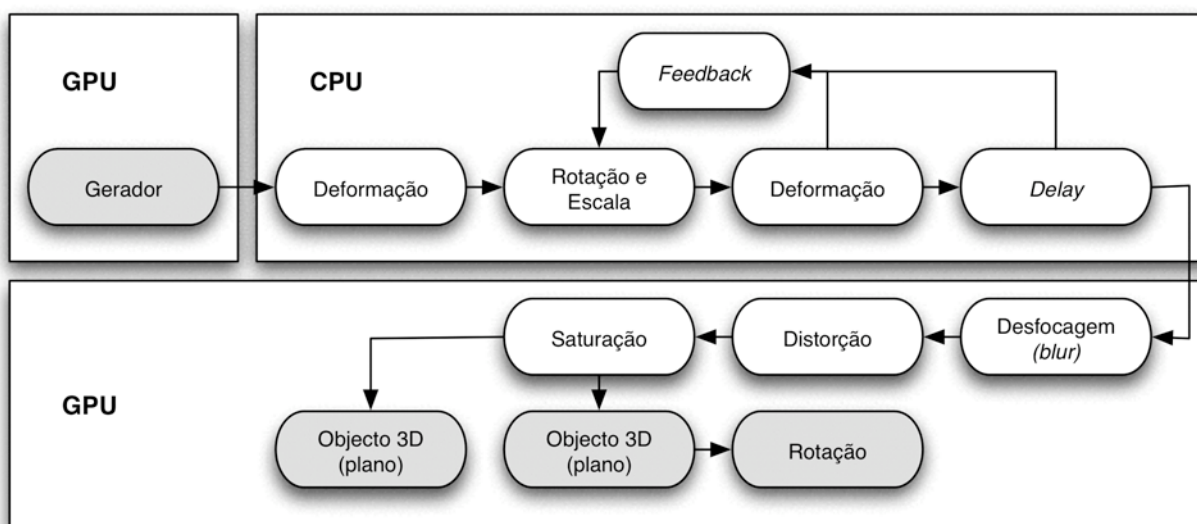


Figura 44 – Estrutura de uma das implementações do Sintetizador Visual

Como base do sistema visual é necessário efectuar o controlo sobre a cadência temporal da geração das imagens. O objecto responsável pelo metrónimo é o [qmetro] (*queue metro*), em vez do comum [metro]. Este permite que seja definido um valor em milissegundos para o envio de cada *trigger*. O [qmetro] envia-os no modo de baixa prioridade, ou seja, utilizando uma “fila de espera”. Se, na altura em que novo *trigger* for emitido a mensagem anterior ainda se encontrar por realizar, esta nova mensagem substitui-a. Desta forma, o [qmetro] é o objecto ideal para um sistema visual em tempo real, onde desejamos obter o número máximo de imagens por segundo, com a possibilidade de diminuição do mesmo sempre que o processamento assim o exija. Cada *trigger* do metrónimo faz accionar um conjunto de instruções de forma sequencial, nomeadamente a eliminação da imagem anterior, o accionamento do processo de desenho, a captura e processamento das imagens resultantes, e a visualização das mesmas.

Os comandos para desenhar e apagar os *frames* são enviados para o objecto [jit.gl.render], responsável pelo cálculo final da cena OpenGL e é responsável pela configuração de várias características da mesma, como a cor da luz ambiente e a sua localização, a localização da câmara virtual, e o ângulo de abertura das lentes virtuais. Outros parâmetros permitem activar funções como o método de iluminação *Gouraud Shading*, a técnica de *anti-aliasing*, ou a possibilidade de mistura de objectos através da definição de transparência.

3.6.2 Áudio Digital

A abordagem de visualização a esta fonte sonora é, como referido anteriormente baseado na forma de onda do próprio sinal áudio. Este elemento gerador é obtido a partir dos objectos [jit.catch~] e o [jit.graph] ou [jit.gl.graph]. O primeiro serve de “ponte” entre o MSP e o Jitter, ou seja, recebe um ou mais sinais MSP e transforma-os em matrizes Jitter. Este objecto possui vários modos de funcionamento relativamente à captura dos sinais, sendo que o modo 3 seleccionado é similar ao comportamento de um comum osciloscópio. A informação relativa às duas matrizes (se a fonte for estereofónica) é separada e enviada para dois objectos [jit.graph], responsável por desenhar os gráficos com os valores obtidos pelo [jit.catch~]. Nos argumentos do objecto são definidas as dimensões do gráfico bem como a sua cor. É possível ainda mudar o modo de visualização como ilustra a Figura 45.



Figura 45 – Os 3 modos de desenho de gráficos do [jit.graph]: linha, contorno preenchido e em espelho

A forma de onda desenhada pelo [jit.graph] contém já informação relativa à amplitude do sinal (eixo Y), mas esta amplitude não é necessariamente sinónimo da nossa percepção de intensidade sonora. Esta depende de outros factores como por exemplo a quantidade de frequências presentes, a não-linearidade do ouvido em relação ao espectro audível, entre outros. Por esta razão resolvemos escalar o valor de amplitude da onda com o valor de *loudness* obtido através do objecto [analyzer~]. Decidimos incorporar ainda a possibilidade de definir um factor de atenuação adicional controlável pelo utilizador.

A selecção da cor é neste caso um parâmetro fundamental na correlação entre som e imagem. Enquanto que em grande parte dos visualizadores existentes a escolha da cor é aleatória ou segue algum algoritmo pré-definido não correlacionado com o áudio, neste caso optou-se por lhe dar alguma relevância.

Partindo dos parâmetros da análise que obtivemos, podemos tentar relacioná-los com os parâmetros de cor de que dispomos. Pareceu-nos mais lógica a definição da cor através do modo HSL (*Hue, Saturation, Luminosity*), permitindo a manipulação independente da matiz, saturação e luminosidade (brilho), do que através do comum modo RGB (Red, Green, Blue). Logicamente, trata-se aqui apenas de uma questão de acessibilidade aos parâmetros de cor, uma vez que continuamos a trabalhar neste modo. Para definir a cor através dos valores HSL, basta enviar uma mensagem do tipo “hsl valor_h valor_s valor_l” para o objecto [swatch], responsável pela selecção de cor RGB. A atribuição dos parâmetros sonoros aos parâmetros de cor foi feita da forma que a seguir descrevemos. As relações estabelecidas entre eles não pretendem ser verdades absolutas, mas julgamos estarem relacionadas, pelo menos numa perspectiva física dos fenómenos sonoro e visual.

Sensação de intensidade (*Loudness*) / Luminosidade

Tanto a intensidade sonora como a luminosidade estão directamente ligadas em termos de fenómeno físico. Ambos definem a “quantidade” de som ou luz, logo será lógico o relacionamento destes dois parâmetros. É de salientar que a utilização do objecto [zmap] é aqui necessária para otimizar o mapeamento destas relações. Se não o fizéssemos, a visualização poderia ficar muito clara uma vez que um conjunto de valores HSL com uma definição de L= 255 daria sempre o branco. Desta forma, optámos por restringir a gama dinâmica deste parâmetro.

Brilho (centro espectral) / Cor (*Hue*)

O centro espectral do sinal áudio, aqui definido como “brilho”, revela-nos, tendo em conta a totalidade das frequências presentes no sinal, qual a frequência média em cada instante. Analisando os respectivos fenómenos físicos, pareceu-nos também lógico associar este parâmetro à selecção da frequência da cor (*hue*), relacionado os sons graves com as cores quentes começando pelo vermelho, e os sons agudos com as cores frias, terminando no violeta. De forma a correctamente relacionar a frequência do centro espectral com a frequência, foi necessário transformar os valores de *brightness* provenientes do [analyzer], definidos em Hz para valores MIDI, compreendidos entre 0 e 127 (objecto [ftom]). A lógica subjacente a esta operação não se prende com os valores MIDI em si mesmo, mas sim com o facto de neste modo a nossa percepção da escala ser linear. Visto que a escala utilizada para a definição dos valores de cor é também linear em termos de percepção, conseguimos obter uma correlação mais efectiva entre estes parâmetros.

O mapeamento a seguir efectuado ([zmap 80 127 0 255]) pode parecer à partida, pelos valores envolvidos, um pouco estranho, mas possui uma justificação simples. Ao analisar os valores obtidos pela leitura de várias músicas, constatámos que estes quase nunca utilizavam os valores inferiores da escala. Isto deve-se ao facto de que qualquer som natural terá à partida uma série de harmónicos, fazendo tender o centro espectral para valores mais elevados. Se o mapeamento não fosse aqui compensado, raramente obteríamos valores mais reduzidos. Sem a restrição efectuada não seria possível obter cores como o vermelho, o laranja, ou o amarelo. Assim, todo o espectro visual é utilizado.

Quantidade de ruído (*Noisiness*) / Saturação

Também estes dois parâmetros estão em termos físicos, e em nossa opinião, em termos de percepção, relacionados. A saturação de determinada cor refere-se à relação entre a frequência principal e as restantes frequências do espectro visível. Uma cor completamente saturada teria teoricamente apenas um único comprimento de onda. Uma cor muito pouco saturada significa que o comprimento de onda principal se evidencia muito pouco relativamente aos restantes. No caso do parâmetro *noisiness* obtido através do objecto [analyzer~], este revela-nos numa escala de 0 a 1, a quantidade de frequências contidas no sinal. Logo, um sinal com um valor baixo implica um conteúdo espectral simples, enquanto que um valor alto revela um sinal

com muito ruído, aproximando-se do ruído branco. Podemos compreender até por esta designação de ruído branco que estes dois parâmetros estão de facto relacionados. No entanto, para correctamente os relacionarmos, devemos ter em atenção que o significado é aqui oposto: um valor baixo de *noisiness* sugere um valor alto de saturação e vice-versa. Por esta razão, utilizámos o objecto [scale] em vez do [zmap] pois este permite-nos inverter a ordem de saída dos valores. Neste caso concreto mapeamos os valores entre 1 e 0 com os valores entre 128 e 255. Não utilizámos a primeira metade da escala (0-127) pois estes valores traduzem-se em cores demasiado neutras, não desejáveis para o aspecto visual em causa.

Outros parâmetros

Para além dos parâmetros acima citados, relacionados com o gerador visual, foram ainda acrescentados módulos de processamento de imagem (modificadores) como a alteração de escala, rotação, desfocagem ou deformação. No entanto, contrariamente aos parâmetros de controlo do gerador, estes não são controlados automaticamente por informações provenientes do sinal áudio. Poderão ser controlados de forma aleatória, pela selecção de programas pré-definidos, ou por controlo manual, utilizando os dispositivos referidos em “Controlo do Sistema Visual” (pág.132). Uma excepção a este princípio foi a associação dos efeitos de deformação da imagem com a detecção de ataque proveniente do objecto de análise ao sinal áudio; de cada vez que um ataque é detectado, é gerada uma matriz que deforma a imagem original com um grau proporcional à intensidade do mesmo.

3.6.3 MIDI

A ideia base para a implementação do sistema visual é significativamente diferente da do conceito anterior. Enquanto que na situação anterior a nossa preocupação incidiu mais na geração e processamento ao nível das texturas, neste caso particular interessou-nos explorar as potencialidades do palco tridimensional. Através da utilização da tecnologia OpenGL torna-se possível alterar em função da música uma série de parâmetros como a cor, a posição e escala dos objectos ou até a própria geometria dos mesmos. Poderemos ainda mapear, à semelhança da situação anterior, texturas que cubram a sua superfície.

Visto que a informação musical proveniente do MIDI consiste num *kit* de bateria/percussão com 8 partes independentes, decidiu-se que cada uma destas seria representada por um objecto tridimensional posicionado de forma distinta no palco visual. O accionar de cada um destes sons provoca a deslocação do respectivo objecto no sentido do espectador (eixo Z), variando o grau conforme a intensidade do ataque.

A programação deste *patch* começou pela recepção dos parâmetros provenientes do sistema sonoro, através de MIDI. Sendo a caixa de ritmos baseada em sons percussivos, interessa-nos apenas o valor de ataque de cada instrumento, proporcionadas pelos parâmetros MIDI de cada *Note On*. O posicionamento dos mesmos

encontra-se já pré-definido. A posição do som na imagem estereofónica poderia ser repercutida no espaço visual. Contudo, no protocolo MIDI os *kits* de bateria/percussão estão afectos a um único canal MIDI, sendo que a mensagem de controlo da panorâmica (*Control Change 10*) afecta todo o canal.¹

Adicionalmente é enviado o valor relativo ao volume geral da caixa de ritmos (*Control Change 7*) afecto ao respectivo canal MIDI. Este último foi por nós relacionado com o nível de transparência dos objectos. Um parâmetro MIDI que merece ainda destaque é a informação relativa ao andamento do sequenciador (*MIDI Clock*). Este parâmetro está a controlar o tempo de deslocação do objecto até à posição inicial, depois deste ter sido accionado; quanto mais lento for o andamento, mais tempo os objectos demorarão até estabilizar. Para que estes objectos só sejam visíveis quando existe conteúdo musical relacionado, à medida que cada objecto se vai deslocando para a “posição de repouso”, o valor do parâmetro de transparência é igualmente afectado, resultando num efeito de *fade-out*. Contudo, este valor terá necessariamente que ser relacionado com o valor de volume geral atrás mencionado. Dado o carácter percussivo dos instrumentos, os objectos aparecem de forma instantânea.

Para além destes parâmetros, são ainda enviados outros para cada uma das 8 instâncias dos objectos [jit.gl.gridshape] como a profundidade inicial, a definição do tamanho dos objectos ou o número de polígonos. Um outro parâmetro enviado consiste numa matriz que funcionará como textura para os objectos. Para além da sua cor própria, cada objecto poderá ainda ser influenciado pelas restantes texturas presentes no palco visual. Demonstrando esta situação, podemos ver na Figura 46 os objectos com uma simples atribuição de cor, à esquerda, enquanto que na imagem da direita os gráficos gerados pela acção do sintetizador sonoro são também utilizados. O mapeamento da textura no objecto é neste caso esférico, de forma a simular a reflexão da área envolvente.

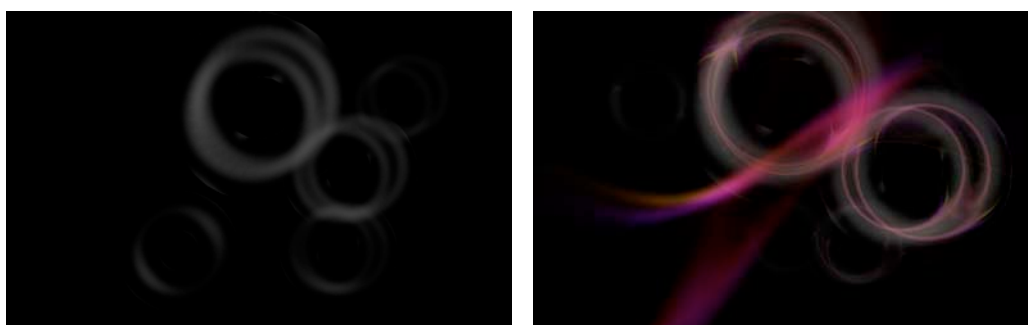


Figura 46 – Objectos gerados relacionados com a caixa de ritmos.

¹ Foi no entanto testada com sucesso esta possibilidade de controlo da “panorâmica visual” através do protocolo OSC.

3.6.4 Síntese Sonora

Uma vez que esta é a abordagem de visualização com maior capacidade de acesso à descrição de eventos musicais e tímbricos, torna-se lógica a decisão de concentrarmos nela os nossos esforços. Assim sendo, esta abordagem é a que possui um maior mapeamento de relações entre os domínios sonoro e visual. De seguida referimos os principais aspectos de programação de relacionamento referentes ao elemento gerador, às cores a ele associadas, e aos vários modificadores, sob a forma de processamento de efeitos.

Gerador

O gerador utilizado nesta última abordagem é similar à do áudio digital e consiste na utilização do objecto [jit.gl.graph] como principal elemento. No entanto, existe uma diferença fundamental entre as duas abordagens: enquanto que na abordagem áudio este objecto é alimentado pelo [jit.catch~], responsável por capturar a forma de onda do sinal áudio, neste caso não existindo essa matéria prima, decidimos construir uma forma de onda de raiz tendo por base uma matriz unidimensional. Para a criação dessa matriz, desenvolvemos duas abordagens distintas: a primeira consistiu em seguir um modelo similar ao utilizado na versão áudio do sistema, baseado na geração de sinais sinusoidais, enquanto que a segunda abordagem consistiu na geração das mesmas através de uma fórmula matemática. De seguida, analisamos cada um destes métodos.

Sinais

A primeira abordagem consistiu em gerar sinais similares aos do sintetizador sonoro, no que diz respeito à frequência e amplitude de cada uma das vozes. Essa informação é disponibilizada através dos parâmetros OSC recebidos. Como referido anteriormente, a onda gerada não reflecte o timbre, ou seja, a estrutura harmónica do som, consiste apenas numa onda sinusoidal com a frequência da fundamental. É de notar que esta decisão foi propositada, visto que a estrutura harmónica pode estar reflectida visualmente através de muitos outros parâmetros.

O problema inerente a esta abordagem é que, ao estarmos a gerar sinais áudio apenas para a representação visual, estamos a consumir poder de processamento útil para o resto do sintetizador visual. Os sinais são processados continuamente a uma frequência de 44100 Hz ou superior, apenas para que seja gerada a sua representação visual. No caso da representação de áudio, referida anteriormente, a utilização dos objectos [jit.catch~] torna-se essencial, mas neste contexto em que a informação já vem “discretizada”, torna-se mais irrelevante a sua utilização.

Expressão Matemática

A segunda abordagem consistiu na utilização do objecto [jit.expr] para a criação das ondas sinusóides (Figura 47), evitando assim a utilização de sinais áudio para esse efeito. Esta opção, depois de optimizada, revelou menos exigências ao nível do processamento, logo foi a opção escolhida e posteriormente desenvolvida para inclusão neste sintetizador visual.

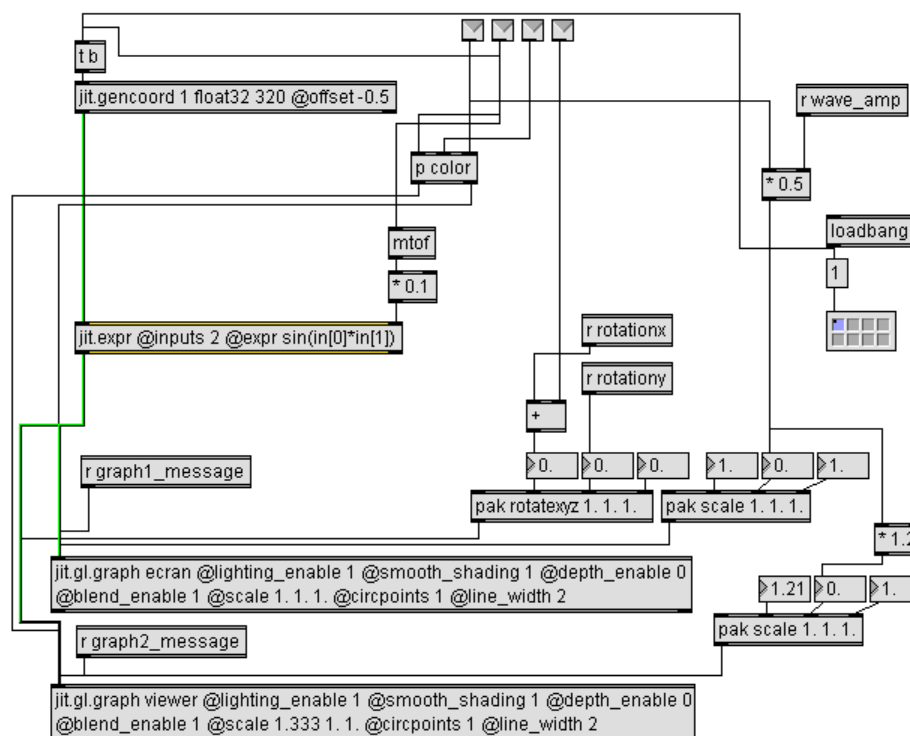


Figura 47 – O patch [glgraph], responsável pela geração das ondas sinusóides através de expressão matemática.

Como inlets temos (por ordem): amplitude, afinação, filtro, rotações.

O objecto [jit.expr] necessita de receber uma matriz com coordenadas para serem por ele processados. Neste caso, o objecto [jit.gencoord] executa essa função. Foi gerada uma matriz unidimensional com 320 valores. Em vez de accionar este objecto através do metrónomo do sistema [qmetro], considerámos ser preferível accioná-lo sempre que existam dados OSC relativos à afinação (nota) a ser recebidos. Assim, este só é accionado quando o valor de afinação da nota é alterada, de forma a poupar recursos de processamento. Visto que a maior parte das vezes a alteração deste valor é pontual (uma excepção será o uso do *pitchbend* no teclado), as vantagens desta abordagem são claras.

O objecto [mtof] converte a informação representada entre 0 e 127 (valores MIDI) em valores de frequência. Os valores MIDI estão definidos numa escala logarítmica, mais relacionada com a nossa percepção, mas neste caso concreto, como pretendemos visualizar a relação geométrica da estrutura harmónica entre as notas a escala deverá ser linear (Figura 48).

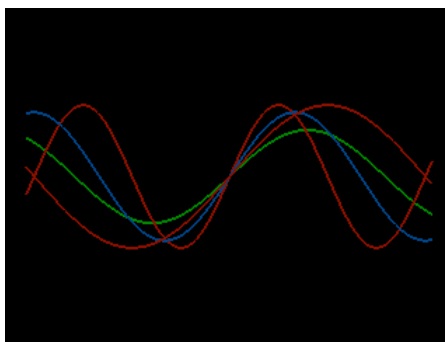


Figura 48 – Aspecto visual de um acorde de Dó Maior

O valor de amplitude sonora não afecta directamente a expressão matemática. Numa primeira abordagem, ao mapear a amplitude desse modo, constatámos que esta opção prejudicaria imenso a *performance* do sistema. Isto porque, contrariamente à afinação, a amplitude não costuma possuir um valor estático, uma vez que é controlada por um envelope ADSR (onde apenas o S é constante) e pode ainda ser modulada por um oscilador de baixa frequência (LFO). Concluimos portanto que o objecto [jit.expr] teria que estar permanentemente a recalcular a onda sinusoidal. A solução por nós encontrada para aliviar o processamento necessário foi a de mapear a amplitude no parâmetro escala (Y) do objecto final OpenGL [jit.gl.graph]. No contexto desta programação é ainda possível controlar a rotação X/Y do objecto gerador. No caso da rotação no eixo Y, esta deriva do módulo de *auto-pan* (panorâmica controlada por um LFO) do sintetizador sonoro.

Cor do Gerador

Para a definição da cor dos gráficos OpenGL decidimos implementar dois modos distintos: o modo correlacionado com o sintetizador, e um outro, onde a selecção de cores é efectuada de forma aleatória. Este último encontra-se aqui presente apenas como termo de comparação, visto que o nosso desejo de exploração de elos de ligação entre o som e a imagem sugere claramente o primeiro modo.

Neste *patch* são definidos os 4 parâmetros essenciais para a selecção da cor dos gráficos a gerar: as cores primárias da teoria aditiva da cor, RGB (*Red*, *Green*, *Blue*), e um canal adicional de definição de transparência, denominado de *alpha* (A). Como foi referido anteriormente, em vez de manipularmos directamente os valores RGB (*Red*, *Green*, *Blue*), é-nos mais intuitivo mapear os vários parâmetros sonoros no modo HSL, por isso optámos por interagir com a definição de cor desta forma. Assim, os quatro parâmetros da cor (HSLA) derivam de quatro parâmetros sonoros distintos. A correlação entre estes foi efectuada segundo os critérios e procedimentos que de seguida expomos.

Matiz

Abordámos já anteriormente algumas relações entre as notas e as cores sugeridas por vários compositores, como Alexander Scriabin, na sua obra orquestral “Poema de Fogo” (1910). Em alguns compositores, estas

associações são sinestésicas, noutras apenas uma racionalização, como no caso de Scriabin, onde se nota uma lógica inspirada no ciclo de quintas. No caso particular deste sistema, não pretendemos chegar a nenhuma conclusão relativamente a esta associação, pois esta revela-se extremamente pessoal. Racionalizar em demasia sobre esta associação não nos parece útil para o trabalho em causa e, muito provavelmente não o será também para quem escute e visualize este sistema. Isto porque a maioria das pessoas não possui o chamado “ouvido absoluto”, ou seja, saber que nota está a ser tocada quando isolado de qualquer termo comparativo. A maioria dos músicos considera muito mais importante possuir um bom ouvido relativo, relacionando melodicamente e harmonicamente as notas num determinado contexto musical. Transpondo isto para o campo visual, parece-nos mais interessante a exploração das relações harmónicas entre as várias cores do que a procura de uma correspondência exacta entre nota e cor, que em nossa opinião não existe.

Para este sistema, à semelhança do modelo criado anteriormente, optámos por definir a cor através da nota tocada. No entanto, neste caso, cada nota da escala possui uma determinada cor, independentemente da oitava utilizada. Sendo assim, a atribuição de cores foi relacionada com as 12 notas da escala musical (Figura 49), e é cíclica, repetindo-se em cada oitava. Esta distribuição teve como único critério a disposição das cores pela ordem do espectro visual, desde os infra-vermelhos até aos ultra-violetas. Para nos abstrairmos das oitavas e mapear os valores de entrada entre 0 e 127 em valores de nota (0 a 11) de forma cíclica, utilizámos o objecto [modulo] ([% 12]), seguido de um escalamento para os valores pretendidos de cor.

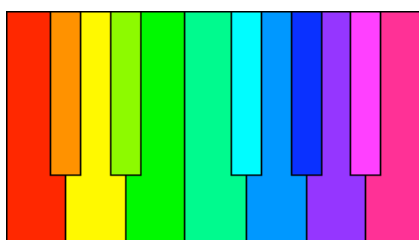


Figura 49 – Mapeamento das diferentes cores na escala de um teclado

Saturação

Como referimos anteriormente, a saturação está relacionada com a pureza da cor, ou seja, a quantidade de cor dominante relativamente às restantes cores do espectro. Face a estas características, pareceu-nos interessante relacionar este parâmetro com a ressonância do filtro do sintetizador sonoro. Essa ressonância pode ser caracterizada por uma certa saturação sonora na área da frequência de corte. Ao contrário dos outros parâmetros sonoros relacionados com a definição da cor, o parâmetro associado à saturação é global, ou seja, igual para todas as notas. Assim sendo, foi possível enviá-lo e recebê-lo através dos objectos [send] e [receive] em vez de *inlets* e *outlets*.

Os únicos ajustes efectuados neste mapeamento foram a inversão dos valores de saída (para que um valor de ressonância baixo se concretize num valor de saturação alta e vice-versa) e a restrição de valores, para que

valores médios de ressonância do filtro continuem a resultar em cores relativamente saturadas. Só quando o efeito é extremo é que o valor de saturação terá um valor mínimo.

Luminosidade

A luminosidade, ou brilho, está aqui directamente relacionada com a frequência de corte do filtro sonoro. Ao caracterizarmos um som, a utilização do termo “brilho” é de alguma forma recorrente. Esta apropriação de palavras ligadas ao domínio visual é justificada pela grande ausência de terminologia própria aplicável à caracterização do mundo sonoro. Contudo, a escolha de adjectivos que descrevam de alguma forma as características sonoras não será arbitrária; ela deriva da relação empírica entre os mundos sonoro e visual que de forma consciente ou inconsciente desenvolvemos. O facto de usarmos a expressão “brilhante” para referir um som com uma grande quantidade de harmónicos, levou-nos a relacionar estes dois parâmetros.

Sendo o filtro do sintetizador sonoro um filtro do tipo LPF (*low-pass filter*), um valor baixo resultará num som com poucos harmónicos; se a frequência de corte chegar a ser inferior à frequência fundamental das notas tocadas, poderemos até deixar de ouvir qualquer som. À medida que aumentamos o valor do filtro, o som possuirá mais componentes harmónicas, aumentando assim a sensação de brilho do mesmo. Para o mapeamento destes parâmetros, necessitámos apenas de utilizar o objecto [zmap] para escalar os valores de entrada, adequando-os aos valores pretendidos para controlo da luminosidade, ou brilho. É de referir que este parâmetro possui valores independentes por nota, uma vez que o valor da frequência de corte varia de acordo com a frequência fundamental das notas tocadas¹.

Canal de transparência (*alpha*)

A definição do nível de transparência é efectuada separadamente dos componentes HSL de cor.

O nível de transparência / opacidade ² está neste caso directamente relacionado com a amplitude da nota tocada e com um factor de multiplicação que advém do valor de mistura; quando o *fader* da mesa de mistura do sistema sonoro estiver em baixo, o valor de transparência é alterado em conformidade. Desta forma podemos obter um equilíbrio visual similar ao equilíbrio sonoro. Sendo nosso intuito que o nível de transparência possa diferir entre os dois objectos [jit.graph] presentes no *sub-patch* anterior, duplicámos as saídas de modo a possibilitar essa independência. No mapeamento efectuado através do objecto [zmap], definimos que a progressão da transparência total (0) até à opacidade total (1) decorre entre os valores 0 e 0.3, sendo que todos os valores acima serão sempre totalmente opacos. Este procedimento serve para que as ondas sinusoidais se possam ser opacas na maior parte dos casos, apenas efectuando um *fadeout* quando a onda decresce consideravelmente com tendência para se extinguir.

¹ Este factor depende da configuração do sintetizador sonoro, nomeadamente do parâmetro de *Key Tracking* do filtro.

² O valor 0 equivale a 100% de transparência, enquanto que um valor de 255 equivale a 100% de opacidade.

Efeitos

Depois de desenhados os objectos geradores, procedemos a uma técnica muito popular em OpenGL que consiste em transformar a cena gerada numa textura (objecto [jit.gl.texture]) de modo a poder efectuar processamento na mesma ao nível dos *pixels* e não da geometria. Podemos assim aplicar uma grande variedade de algoritmos de efeitos implementados na forma de *fragment shaders* programáveis (GLSL), já referidos anteriormente e disponíveis através do objecto [jit.gl.slabs]. De seguida abordamos de forma resumida os principais efeitos utilizados e suas particularidades de implementação.

Realimentação

Pela análise efectuada anteriormente a outros visualizadores de áudio (Cap. 2.6, pág. 66), rapidamente podemos constatar que processo de *feedback* é um dos efeitos mais comuns neste domínio da visualização. Através dele, podemos conferir um determinado grau de persistência visual no tempo. Segundo Michel Chion [6], a nossa forma de ouvir não é contínua, mas sim dividida em breves segmentos, sendo que o que memorizamos e percebemos a cada instante são os últimos 2 ou 3 segundos do som à medida que este se desenvolve. A utilização deste processo poderá servir para simular essa mesma persistência e, quando aliado a outros factores intermédios de modificação, como os referidos abaixo, permite a obtenção de efeitos visuais bastante mais sofisticados. A imagem realimentada necessita de ser alvo de um determinado factor de atenuação pois se assim não fosse o efeito de *feedback* seria permanente. Decidimos controlar este factor de multiplicação que regula a quantidade de realimentação do sistema, e por consequência directa, o tempo do efeito, através do sistema sonoro. Dadas as suas características de fazer perdurar no tempo eventos já ocorridos, achámos interessante a sua correlação com o tempo de reverberação do processador de efeitos existente no sistema sonoro. O objecto [zmap] foi aqui utilizado para acertar a duração dos dois parâmetros, para que, por exemplo, um tempo de reverberação de 2 segundos resulte sensivelmente em igual tempo de *feedback*. É no entanto importante referir que a duração deste efeito não depende exclusivamente deste factor de multiplicação, mas também, como é lógico, do próprio *framerate* da visualização. Sendo este *framerate* variável conforme o processamento a ser efectuído, num desenvolvimento mais extensivo deste sistema seria necessário contemplar também este factor para a definição correcta do valor de *feedback*.

Rotação e escala

Como já referido, se não existisse nenhum processamento adicional, pouco interesse estético teria o efeito de *feedback*. Utilizámos por isso um *shader* GLSL que nos permite processar a matriz em termos de rotação e de escala. Este objecto revelou-se extremamente interessante pelas possibilidades que revelou em conjunto com a técnica de *feedback*. A controlar os vários parâmetros a ele associados poderemos ter a mesa digitalizadora Wacom ou o *gamepad* e o próprio sintetizador sonoro. A mesa digitalizadora controla a escala nos eixos X e Y (valores entre 0.9 e 1.1, sendo que o valor 1 representa o tamanho original). Podemos

observar no primeiro quadro da Figura 50 os efeitos de um escalonamento de 0.9 e no segundo quadro de 1.1 (escalonamento igual para os eixos X e Y). No terceiro existe a particularidade do escalonamento ser diferenciado para os dois eixos ($X=0.9$, $Y=1$).

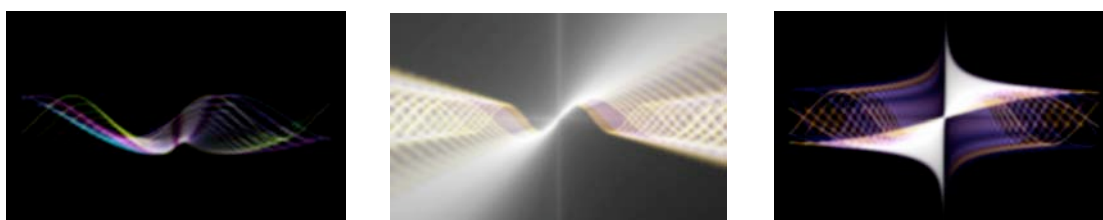


Figura 50 – Diferentes efeitos visuais proporcionados pelo processamento de *feedback* aliado à escala

Em todos os casos acima descritos, o posicionamento da âncora de escala e rotação foi o centro visual. Achámos interessante relacionar a modulação cíclica (LFOs) da amplitude e frequência de corte do filtro com o posicionamento desta âncora. Assim sendo, relacionámos a variação de amplitude com o posicionamento da âncora no eixo X, enquanto que a frequência de corte está relacionada com o eixo Y. Destes dois mapeamentos, o que melhor resulta é em nossa opinião este último, visto que as linhas sinusoidais que servem de base para a construção visual possuem à partida um alinhamento horizontal. No entanto, se existir alguma rotação (Figura 51) o efeito visual do posicionamento da âncora no eixo X pode ser também bastante perceptível. A quantidade de rotação pode ser controlada a partir de dispositivos adicionais como a caneta da mesa digitalizadora ou o *gamepad*.



Figura 51 - Diferentes efeitos visuais proporcionados pelo processamento de *feedback* aliado à alteração da âncora (esquerda), da rotação (centro) e de ambas (direita)

Deformação

Este efeito é responsável por deformar a matriz de entrada através de um efeito de reposicionamento (Figura 52). Quando este se encontra activo, possui dois modos de funcionamento distintos: pode ser accionado sempre que existe o accionamento de uma nota musical (sendo a amplitude do efeito proporcional ao ataque da mesma), pode ser accionada periodicamente, segundo a recepção de *andamento* musical proveniente do protocolo OSC. Alternativamente, poderá ser accionado o controlo manual por parte de um instrumentista visual. É conveniente reforçar a ideia de que o aspecto final deste efeito é fortemente influenciado pelo processo de *feedback*.

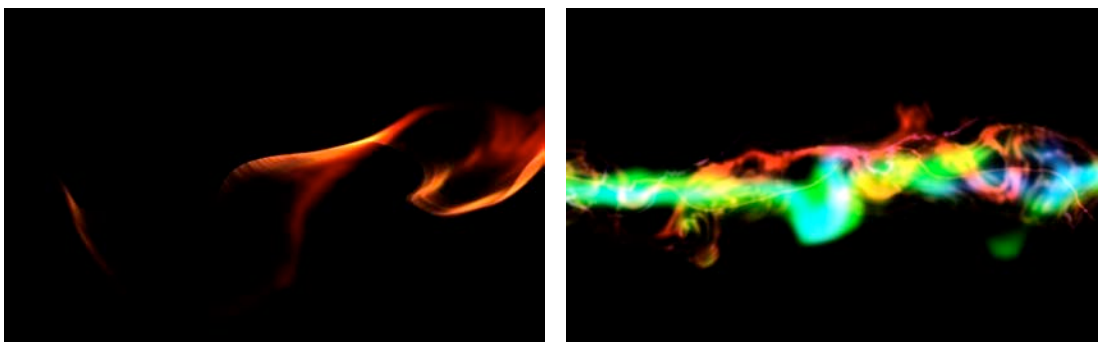


Figura 52 – Efeito de deformação de imagem associado a um processo de *feedback*

Desfocagem (*blur*)

O efeito de desfocagem aqui utilizado é obtido através do processamento em 2 passagens distintas. São elas a desfocagem no eixo horizontal, seguida de uma desfocagem no eixo vertical. O parâmetro de amplitude de desfocagem é controlado pelo sistema sonoro e foi por nós relacionado com a frequência de corte do filtro (Figura 53). Considerámos existir aqui uma relação interessante entre estes dois parâmetros. A quantidade de focagem / desfocagem pode estar directamente relacionada com um filtro LPF no domínio da imagem, logo achámos oportuna o mapeamento destes dois parâmetros. Ao aumentarmos o valor do filtro, o som resultante irá conter mais componentes de alta frequência, ou seja, podemos considerar que o som será mais pormenorizado. Neste caso, a imagem deverá estar pouco desfocada, de forma a manter os pormenores de alta frequência espacial. Se o valor do filtro for reduzido, o som resultante conterá apenas componentes de baixa frequência, resultando visualmente numa maior desfocagem (menor pormenor visual). Esta relação é realçada pelo facto de culturalmente termos a tendência de caracterizar um som com uma extensa estrutura harmónica como “áspero”, “rugoso” ou “pontiado”, e um som com poucas componentes harmónicas como “suave”, “macio” ou “redondo”.

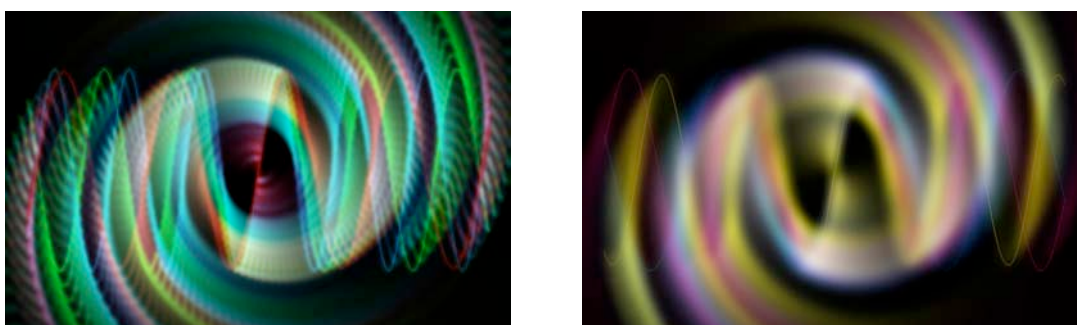


Figura 53 – Parâmetro de desfocagem com valor baixo (esquerda) e alto (direita)

Distorção radial

Este *shader* efectua uma distorção radial com variação sinusoidal. Em termos de resultado visual, dá-nos a sensação de um mapa de relevo ondulado aplicado ao plano OpenGL final (Figura 54). Esta distorção contempla vários parâmetros interessantes, em particular, a frequência e amplitude da distorção, e

posicionamento no eixo horizontal e vertical do centro de distorção. Este efeito poderá ser controlado pela caneta da mesa digitalizadora; a posição da caneta na mesa define a origem da distorção (X,Y) e a pressão executada controla a quantidade de distorção, ou seja, o parâmetro de amplitude. Este efeito poderá também ser controlado pelo sintetizador sonoro. A associação sugerida neste caso foi com o processador de *flanger*. O efeito *flanger* altera as características tímbricas de forma cíclica. Sendo a sua contribuição para a sonoridade final bastante significativa, pareceu-nos interessante efectuar este relacionamento, visto que também este efeito de deformação visual partilha essas mesmas características. Os parâmetros deste efeito relacionados com o sintetizador sonoro foram a amplitude e a frequência. A amplitude está relacionada com a quantidade de efeito do processador, ou seja, a relação entre o som original e o som processado. No caso da frequência, este parâmetro está associado a um oscilador de baixa frequência, responsável pela sua modulação cíclica.



Figura 54 – Diferentes configurações do efeito de distorção

Saturação da cor

O *shader* utilizado para este efeito denomina-se “*sharpen*”. É aqui utilizado não para tornar mais nítido o aspecto visual, mas para saturar as cores e aumentar o contraste. O efeito *sharpen* tenta realçar as linhas das imagens aumentando o contraste e a saturação, tornando as passagens de cor menos progressivas (Figura 55). Todo este processo distorce a cor e a própria forma original da imagem, transformando-a numa imagem mais “agressiva”. Este efeito tem apenas um parâmetro de controlo associado, de quantificação do processamento, que neste caso é definido pelo sintetizador sonoro. Dado o aspecto visual deste efeito, achámos conveniente associá-lo ao processamento de distorção do sintetizador sonoro, visto que ambos fornecem o mesmo carácter “agressivo” ao objecto processado.

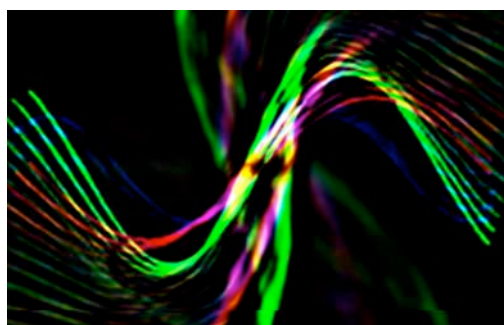


Figura 55 – Efeito provocado pelo *shader* “sharpen”

Planos OpenGL

Antes de ser visualizada, à textura resultante destes processos é acrescentada a informação relativa ao canal de transparência. Essa informação deriva de um *shader* que analisa a textura de forma a obter os valores da luminosidade geral dos restantes canais (RGB), e grava o resultado no canal de transparência (*alpha*). Esta nova textura pode posteriormente ser mapeada nos objectos OpenGL presentes na cena, sejam eles planos simples ou objectos com uma geometria mais complexa. Através da definição do canal de transparência é assim possível sobrepor planos visualmente e ser capaz de os visualizar em simultâneo.

Uma das motivações originais para a existência de vários planos com a mesma textura associada foi a de poder relacionar o sistema visual com o efeito *chorus* do sintetizador sonoro. Este efeito consiste basicamente na duplicação do sinal e aplicação de uma ligeira desafinação a cada um, de forma simétrica, e posteriormente enviar cada sinal para um canal distinto (canal esquerdo e direito). O resultado sonoro deste efeito é tornar o som mais cheio. Inspirados nesta técnica, optámos por controlar o desfasamento horizontal dos 2 planos com a quantidade de efeito *chorus* definida no sintetizador sonoro (Figura 56). Depois de mapeados os valores de forma a se adequarem à deslocação em causa enviámos esse valor para o primeiro plano. Para que o segundo plano se movimente de forma simétrica, utilizámos o valor invertido.

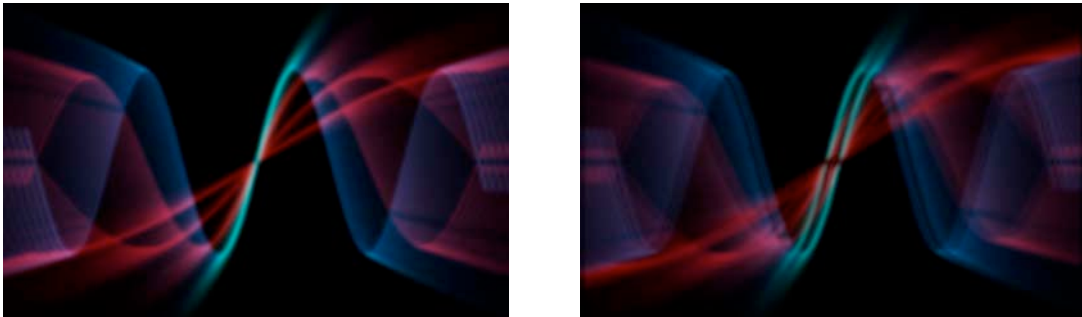
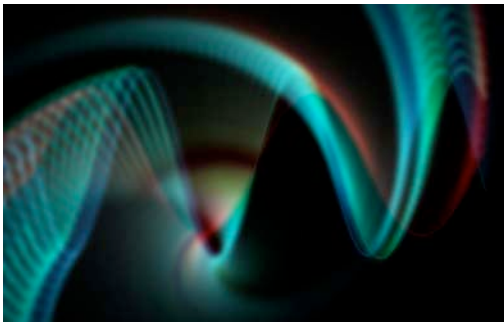
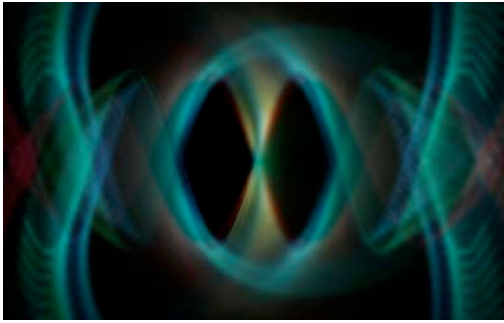

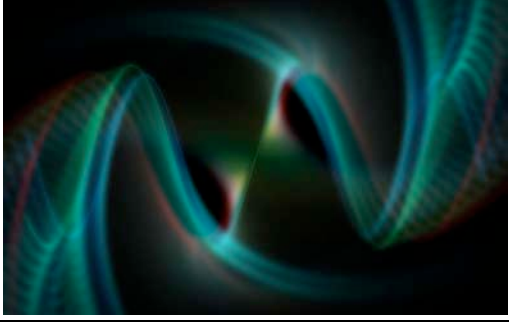


Figura 56 – Imagem normal (esquerda) e imagem com desfasamento horizontal aplicado (direita)

Depois de efectuado este mapeamento, decidimos explorar as possibilidades referentes à existência dos dois planos através da rotação do segundo plano. A relação entre os dois planos torna-se assim simétrica. Foram programados 4 estados distintos, como mostra a Tabela 9.

Tabela 9 – Diferentes estados de rotação do segundo plano OpenGL

Estado	Aspecto visual
Normal (0 0 0)	 A 3D visualization showing a complex, wavy, and intertwined structure. The structure is composed of multiple layers or surfaces that are colored with a gradient from dark blue to bright cyan and green, with some red highlights. The overall shape is elongated and has a sense of depth and movement.
Simetria no eixo X (180 0 0)	 A 3D visualization showing a structure that is symmetric about the X-axis. The structure consists of two main lobes or wings that are mirrored across a central vertical plane. The colors are primarily cyan and green, with some red highlights at the tips of the lobes.
Simetria no eixo Y (0 180 0)	 A 3D visualization showing a structure that is symmetric about the Y-axis. The structure is composed of two main lobes or wings that are mirrored across a central vertical plane. The colors are primarily cyan and green, with some red highlights at the tips of the lobes.
Simetria no eixo Z (0 0 180)	 A 3D visualization showing a structure that is symmetric about the Z-axis. The structure is composed of two main lobes or wings that are mirrored across a central vertical plane. The colors are primarily cyan and green, with some red highlights at the tips of the lobes.

3.7 Métodos de Controlo

O controlo do sistema audiovisual proposto constitui uma das preocupações mais relevantes num contexto muito particular de *performance*. A decisão de que parâmetros expor ao utilizador e de que modo o fazer é o objecto de análise desta secção. Numa primeira fase, dada a simplicidade das abordagens que partem do áudio e do MIDI como fonte sonora comparativamente à terceira abordagem sugerida, que tem como fonte de informação a própria estrutura de produção sonora, decidimos concentrar a nossa atenção nesta última. Numa segunda fase, debruçamo-nos sobre o controlo do sistema visual, em particular na sua relação com o sintetizador sonoro atrás citado.

3.7.1 Controlo do Sintetizador Sonoro

O controlo do sintetizador sonoro é um aspecto fundamental de todo o sistema por nós proposto, uma vez que este, para além de gerar o som, funciona também como a principal forma de controlo do sintetizador visual. Estando o sistema idealizado destinado a ser controlado principalmente pelos músicos, torna-se assim ainda mais importante que o método de controlo do sintetizador seja versátil, permitindo uma grande manipulação musical e tímbrica, resultando também numa grande variedade de resultados visuais.

Num trabalho com um carácter mais exploratório, poderíamos “aventurarmo-nos” na procura do *interface* ideal para o controlo do sistema proposto mas, sendo esta tese norteadada pela sua possibilidade de aplicação a casos concretos e comuns na indústria, decidimos ser mais “conservadores”, tendo elegido dois métodos de controlo possíveis: um teclado MIDI¹ com a particularidade de possuir vários métodos de controlo adicionais como *faders* e botões rotativos, e o próprio *interface* gráfico da aplicação utilizada para construir o sintetizador sonoro, o Reaktor.

É possível ainda acrescentar um sequenciador MIDI ao sistema, de forma a potenciá-lo para outras tarefas menos adequadas à *performance* em tempo-real (Figura 57). Esta abordagem, graças às funcionalidades de gravação, edição e processamento de informação MIDI, permite construir a *performance* ideal em termos audiovisuais. Estando o sequenciador e o sintetizador sonoro a correr em máquinas diferentes, poderão ser interligados através de MIDI. De forma a não ser necessária a utilização dos tradicionais cabos e respectivos *interfaces* MIDI, poder-se-á recorrer à comunicação da informação MIDI através de uma rede local baseada

¹ Apesar do principal método de interacção elegido ter sido o teclado, este poderia ser substituído por qualquer outro instrumento musical compatível com o protocolo MIDI sem grandes necessidades de configuração adicional.

na arquitectura protocolar TCP/IP. Estando o sequenciador e o sintetizador sonoro residentes na mesma máquina, é também possível partilhar informação MIDI entre estas aplicações.¹

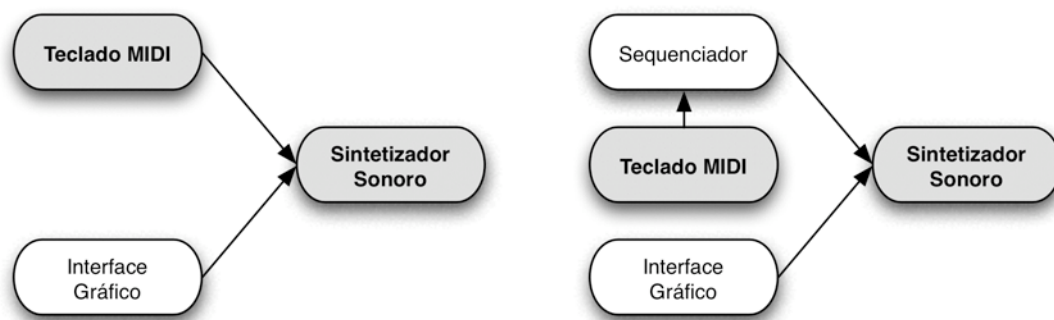


Figura 57 –Possibilidades de controlo do sintetizador sonoro

Controlador MIDI Edirol PCR-M50

O controlador MIDI utilizado na elaboração desta tese foi o PCR-M50 da Edirol (Figura 58). Este teclado possui 49 teclas, um *pitchbend/modulator* e 27 controladores completamente programáveis: 8 botões rotativos, 8 *faders*, 9 botões (*on/off*) e a possibilidade de ligação de 2 pedais externos. Além de possuir o tradicional protocolo MIDI para ligação a outros equipamentos e/ou computadores, possui ainda o protocolo USB, permitindo a ligação ao computador de forma directa, evitando a utilização adicional de um *interface* MIDI.



Figura 58 – Controlador MIDI utilizado

As características de um controlador MIDI como este são extremamente interessantes na medida em que permitem um controlo da *performance* musical muito mais completo, possibilitando para além do tradicional teclado, o controlo de parâmetros sonoros adicionais como a frequência de corte de um filtro, a sua ressonância, o nível de determinado efeito ou a definição do contorno de um *envelope*.

Uma limitação de qualquer controlador MIDI assenta no facto da resolução dos seus controladores ser algo reduzida quando aplicados a determinadas parâmetros. Como foi abordado anteriormente, quase todos os parâmetros MIDI possuem uma resolução de 7 bit, ou seja, 128 possibilidades de quantificação distintas. Se,

¹ Para mais informações sobre estas possibilidades, consultar o capítulo 3.2.1, “Transmissão de MIDI e Áudio entre aplicações locais e em rede”, pág. 80

aplicado à música de uma forma geral, essa limitação não será na maior parte dos casos significativa (128 notas possíveis, 128 intensidades, etc.), no caso do controlo dos vários parâmetros do sintetizador sonoro e visual, poderá ser um pouco mais notória. Dependendo do tipo de parâmetro a controlar e do seu grau de variação, pode ser necessária a restrição dos valores “possíveis” a um subconjunto de valores considerados mais “úteis”, tendo em vista o mapeamento MIDI do parâmetro em causa.

É de salientar que o controlador MIDI utilizado possui a facilidade de armazenamento de memórias com a configuração de todos os parâmetros do teclado, como a atribuição de mensagens MIDI a cada um dos controladores. Assim, torna-se possível, por exemplo, construir várias memórias destinadas a controlar diferentes instrumentos musicais ou até mesmo o sistema visual de forma directa.

Interface gráfico do Reaktor

O Reaktor possui dois ambientes de funcionamento, denominados por *Structure* e *Panel*. O primeiro será o indicado para a programação dos equipamentos, enquanto que o último será o eleito para a operacionalidade dos mesmos. A este nível, o Reaktor distancia-se de outros ambientes como o Max/MSP, ou o Pd, pois estes, apesar de possibilitarem a construção de *interfaces* gráficos, não o fazem com a mesma facilidade. Podemos ver na Figura 59 o aspecto visual do Sintetizador por nós construído na aplicação Reaktor.

Um dos problemas inerentes à operacionalidade de um sintetizador será o nível de acesso aos parâmetros que é proporcionado pelo *interface* e a sua facilidade de utilização. Estes dois aspectos serão, à partida algo opostos: a acessibilidade a demasiados parâmetros torna o sistema mais complexo, resultando numa menor facilidade de utilização. Por isso, um compromisso entre estes é essencial para a concepção do painel de programação de um sintetizador. No nosso caso, optámos por não incluir alguns controlos possíveis mas considerados irrelevantes para a maior parte das situações de utilização, delimitámos e legendámos cada módulo para uma mais rápida compreensão e tentámos agrupar visualmente módulos com funções idênticas, como é o caso do *LFO1* e *LFO2* ou do *Amplitude-Env* e *Filter-Env*.

Em termos de funcionalidade permitida pelo *interface*, é importante referir a possibilidade de registo de todas as configurações dos instrumentos através dos chamados *presets* ou *snapshots*. Existe ainda uma funcionalidade que no contexto em que esta tese se insere pode ser muito interessante: podemos configurar dois modos de visualização distintos, A e B (canto superior esquerdo de cada *instrument*). Um deles pode ser utilizado para a visualização de todos os controlos do sintetizador sonoro e o outro poderá ser aproveitado para restringir a visualização apenas aos controlos que se encontrem relacionados com o sintetizador visual. Assim, pela análise do painel, o utilizador poderá ter uma visão mais clara dos elementos de controlo que afectam de alguma forma o sistema visual.

Na Figura 60 podemos observar a mesa de mistura, por nós adaptada e os processadores de efeitos *Dual Sync* *Delay* e *Space Master 2*, *instruments* provenientes da biblioteca que vem fornecida com o Reaktor. Este último mostra que é possível desenvolver *interfaces* mais sofisticados e apelativos com esta aplicação, através da importação de imagens.



Figura 59 – Pannel do Sintetizador criado no Reaktor



Figura 60 – Pannel da Mesa de Mistura, do processador de efeitos *Space Master 2* e *Dual Sync Delay*

3.7.2 Controlo do Sistema Visual

Tendo esta tese o intuito de permitir uma visualização artística de uma *performance* musical, fará todo o sentido que a maior parte dos parâmetros de controlo do sistema visual sejam provenientes dessa mesma *performance*, sejam estes obtidos a partir do áudio digital, da informação MIDI ou mesmo da estrutura interna de um sintetizador. O problema inerente a esta abordagem anterior é a sua total dependência nas possibilidades tecnológicas de decifrar o conteúdo musical. Se em parâmetros como a natureza espectral (áudio), a descrição das notas (MIDI) ou a frequência de um LFO (síntese), esta abordagem se revela suficiente, a necessidade de uma análise semântica a um nível mais alto, capaz de descrever a estrutura musical ou o discurso melódico, torna-se consideravelmente mais difícil de satisfazer. Assim sendo, numa situação de *performance*, poderemos ter aliado ao músico instrumentista um outro interveniente humano que influencie activamente o resultado visual. Os parâmetros de baixo nível atrás referidos poderão controlar de forma automática determinados parâmetros do sintetizador visual, enquanto que o controlo manual por parte deste segundo elemento potenciará a descrição de informações musicais de um nível mais elevado, como os referidos aspectos composicionais (Figura 61). A introdução deste “instrumentista” visual no sistema constitui assim com toda a certeza uma mais valia para o resultado final da *performance* audiovisual. Neste trabalho, incluímos suporte para várias formas de controlo manual directo do sistema visual: o tradicional rato e teclado, *interface* gráfica, uma mesa digitalizadora Wacom, um *gamepad* da Logitech, controladores MIDI dedicados e sequenciadores (Figura 62).

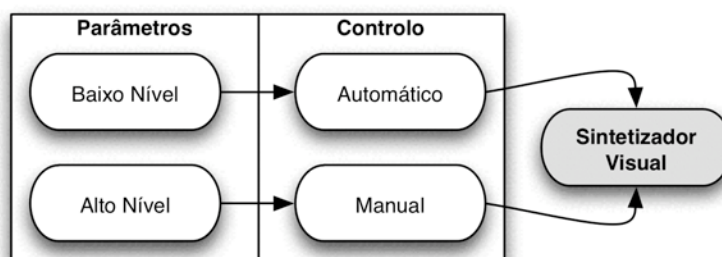


Figura 61 – Controlo partilhado do sintetizador visual

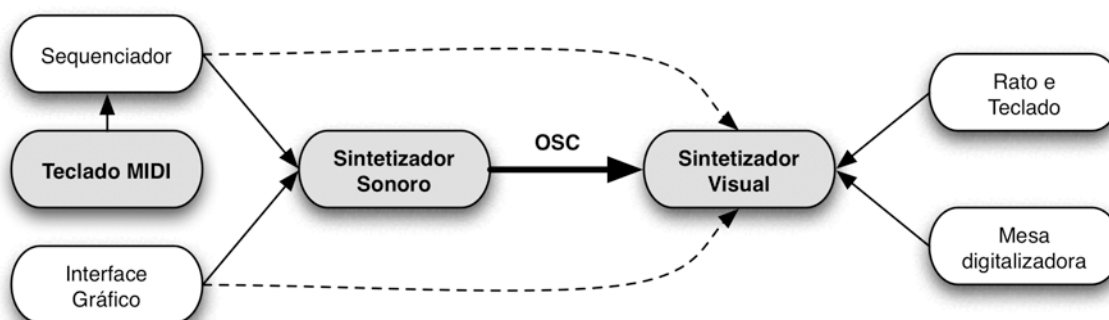


Figura 62 – Exemplo de possibilidades de controlo do sintetizador visual

Rato e Teclado

É possível controlar a navegação no espaço tridimensional do sintetizador visual com o rato e teclado, função proporcionada pelo objecto [jit.gl.handle]. Torna-se assim possível rodar a cena nos 3 eixos, ao clicar e movimentar o rato. Ao utilizar a tecla “Control”, poderemos movimentar-nos na cena nos eixos X e Y, e utilizando a tecla “Alt” poderemos fazê-lo no eixo Z (profundidade). Foram ainda mapeadas teclas que permitem reverter ao reposicionamento e alinhamento original da cena, bem como a comutação da função de rotação automática.

O teclado do computador associado ao sintetizador visual pode ser utilizado para a activação e desactivação de parâmetros de variação não progressiva, ou seja, do tipo ligado/desligado ou por incrementos. Este facto não sugere que as funções controladas por este método sejam necessariamente simples. Continua a ser possível accionar parâmetros mais complexos ou até um determinado conjunto de parâmetros pré-programados. No Max/MSP não foi criado nenhum *sub-patch* que englobasse todas as funções controladas pelo teclado, visto que seria mais fastidioso estar a criar múltiplos [send] e [receive] só para o efeito. Desta forma, achou-se mais conveniente a colocação dos objectos [key] e [select] junto das funções a controlar, em cada um dos *patches* criados. Algumas das funções de controlo atribuídas ao teclado foram a activação/desactivação do modo *fullscreen*, da geração de imagem e do áudio, o controlo de simetria de planos OpenGL, a activação de efeitos visuais e de *presets*.

Mesa digitalizadora

A mesa digitalizadora utilizada, a Wacom Graphire 2, é uma superfície de controlo popular pela sua acessibilidade em termos de custo e de operacionalidade. Ao contrário de um comum controlador MIDI, com uma resolução limitada (128 valores), este tipo de mesas possui uma resolução bastante mais alta, quer em termos espaciais, permitindo controlar parâmetros que necessitem de maior rigor. Em termos ergonómicos, a utilização da forma tradicional de uma caneta torna o manuseamento extremamente intuitivo. As potencialidades de controlo destes sistemas é elevada, ao permitir o envio de um grande conjunto de parâmetros, como o posicionamento nos eixos “X” e “Y”, o nível de pressão efectuado (512 níveis), o azimuth, a altitude, o grau de rotação da caneta relativamente à mesa e a activação de botões.

Para podermos tirar partido deste controlador no ambiente Max/MSP, foi necessária a inclusão de um *external* desenvolvido por Olaf Matthes para a versão Windows¹ do Max/MSP. Com este objecto, torna-se possível aceder a qualquer mesa digitalizadora que possua *drivers* compatíveis com a *library WinTab*. A Wacom Graphire2 utilizada nesta tese suporta o envio dos seguintes parâmetros: posição X, posição Y e nível de pressão (tipo *float*, entre 0 e 1), e botões (tipo *int*, valor 0 ou 1). Possui ainda modos de funcionamento distintos: o modo de caneta e o modo de borracha, quando a caneta é invertida. Utilizámos

¹ <http://www.akustische-kunst.org/maxmsp/>

esta diferenciação para poder duplicar as possibilidades de controlo dos restantes parâmetros. Foi contemplada ainda a possibilidade de, através do pressionar de uma tecla, momentaneamente suspender o envio de informação de pressão relativo à caneta / borracha. Esta funcionalidade é útil, por exemplo, quando desejamos manter um determinado valor de pressão de forma permanente, podendo retirar a caneta da superfície. Através das programações efectuadas, no total ficámos com a possibilidade de poder controlar 11 parâmetros distintos através da Wacom. Na Tabela 10 podemos ver os parâmetros eleitos para serem controlados através deste equipamento.

Tabela 10 – Parâmetros atribuídos a cada função da mesa digitalizadora

Modo	Função	Nome utilizado no Max ([s] e [r])	Parâmetro controlado
Modo Caneta	Posição X	1wacomx	Valor do parâmetro de escalamento da textura no eixo X
	Posição Y	1wacomy	Valor do parâmetro de escalamento da textura no eixo Y
	Pressão	1wacomp	Valor de rotação
	Botão 1	1wacom1	
	Botão 2	1wacom2	
	Botão 3	1wacom3	Inversão do sentido de rotação
Modo Borracha	Posição X	2wacomx	Posição no eixo X do centro da Deformação
	Posição Y	2wacomy	Posição no eixo Y do centro da Deformação
	Pressão	2wacomp	Quantidade de Deformação
	Botão 1	2wacom1	
	Botão 2	2wacom2	

GamePad

A utilização de um *gamepad* permite um elevado grau de controlo do sistema visual, graças ao número de botões e de *joysticks* analógicos e digitais. Também o baixo custo destes dispositivos, dada a sua grande implementação no mercado, torna esta solução bastante atractiva. O *gamepad* por nós utilizado foi o modelo RumblePad 2, da Logitech (Figura 63). Este possui as seguintes possibilidades de controlo: 12 botões (*on/off*), 2 *joysticks* analógicos (valores X/Y progressivos) e 1 *joystick* digital (8 valores posicionais distintos).



Figura 63 – Gamepad utilizado

Contrariamente à mesa digitalizadora, neste caso não recorremos a nenhum *external* adicional para inclusão deste dispositivo no ambiente Max/MSP/Jitter. Utilizámos o objecto [hi] (*Human Interface*), para a recepção dos seus valores. Sendo este um objecto genérico para qualquer dispositivo de entrada, foi necessária a análise do tipo de valores obtidos para cada função e consequente programação de um *sub-patch* de configuração do dispositivo, de forma a abstrair a programação do sistema visual desta necessidade (Figura 64).

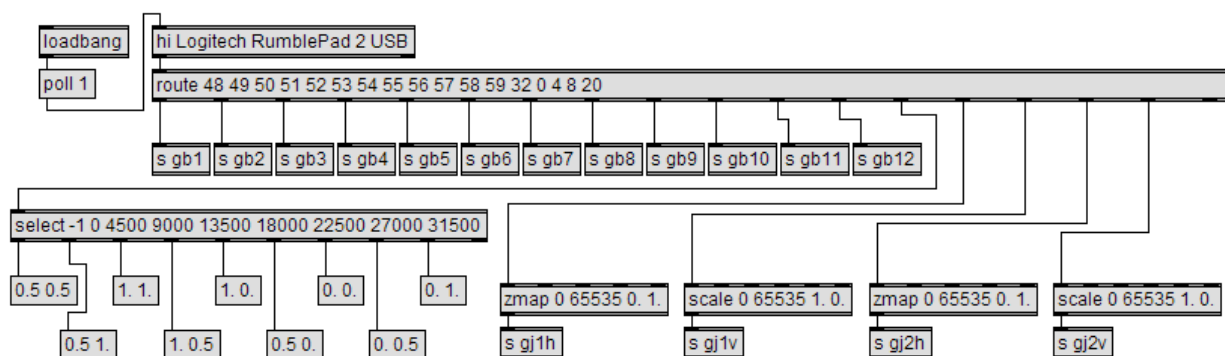


Figura 64 – Patch responsável pela configuração e envio das mensagens do *gamepad* para o sistema visual

Controlador MIDI e Sequenciadores

Pretende-se que o sistema visual proposto possa ser controlado através do protocolo MIDI. Para isso foi criado um *patch* no Max/MSP que recebe as mensagens de Control Change e relaciona-as com os parâmetros do sistema visual. Desta forma, torna-se possível ligar um controlador MIDI equipado com *faders* e/ou botões rotativos directamente ao computador responsável pelo sistema visual e, num contexto de *performance* em tempo-real controlar determinadas características visuais.

Em todas as abordagens anteriores partimos de uma premissa fundamental: a *performance* em tempo-real. Contudo, a visualização criativa da música não se encontra restrita apenas a esta realidade. Quando o propósito da visualização for, por exemplo, o de construir um vídeo musical, será útil uma outra abordagem, não confinada no acto da sua construção às restrições temporais que a *performance* em tempo-real exige. Assim sendo, e visto que o visualizador proposto pode ser controlado através do protocolo MIDI, é possível utilizar uma tradicional aplicação de sequenciação MIDI como o Cubase, o Logic, ou o Digital Performer para gravar e editar com toda o cuidado necessário os parâmetros de controlo ao longo do tempo. Para que as informações musical e de controlo estejam síncronas, será conveniente que o próprio sequenciador esteja síncrono com a reprodução musical (ver Figura 65). Caso a sequenciação musical e visual estejam residentes em máquinas ou aplicações distintas, poder-se-á utilizar o protocolo MTC (MIDI Time Code) para sincronizar os dois sistemas.

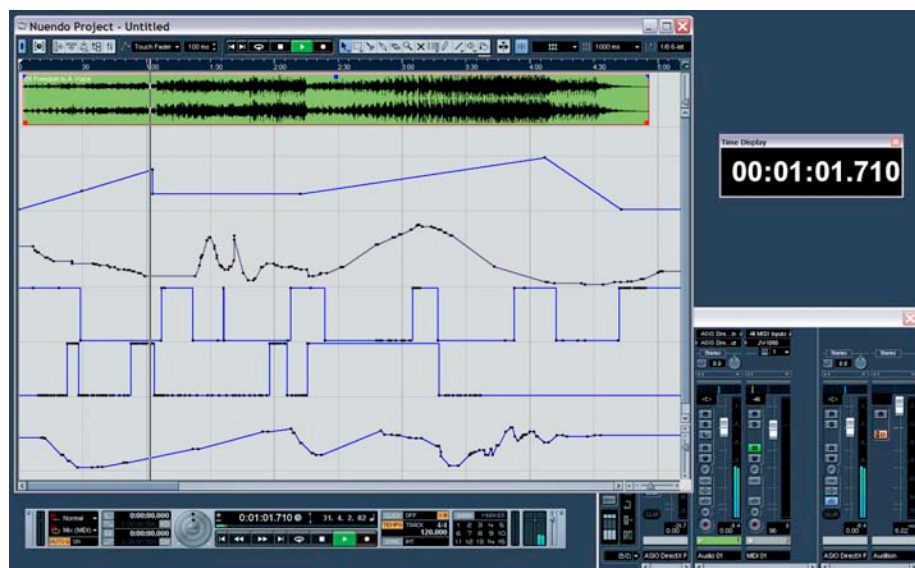


Figura 65 – Sequenciador Áudio/MIDI com a programação de parâmetros de controlo do sintetizador visual

Para que o sistema proposto possa funcionar de uma forma mais completa, será desejável que se possam armazenar, editar e reproduzir no sequenciador não só os dados de controlo originários de controladores MIDI, mas também provenientes dos restantes dispositivos (OSC, mesa digitalizadora, *gamepad*, etc.). Torna-se assim necessária a programação de um estágio intermédio de conversão destes dados (entre a recepção dos mesmos no ambiente Max/MSP e a sua associação aos parâmetros de controlo visual para um formato capaz de ser mapeado para a linguagem MIDI (Figura 66). Pelas suas características, o tipo de mensagem MIDI mais apropriada a este propósito é o *Control Change*, pelo que todos os parâmetros desejados necessitam de ser escalados para valores inteiros contidos entre 0 e 127 ou, alternativamente, em casos em que os parâmetros necessitem de uma resolução acrescida, poderão utilizar-se pares de controladores (MSB/LSB), permitindo a especificação de 16.384 valores (14 bit).

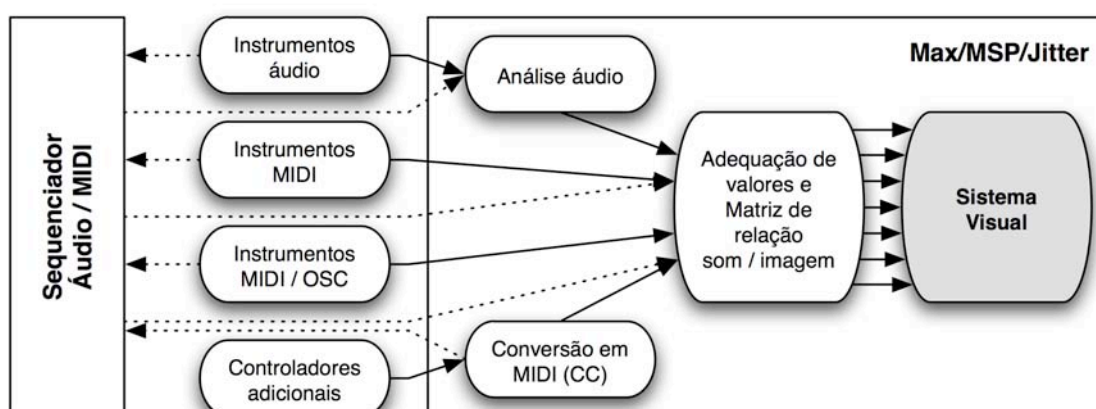


Figura 66 – Possibilidade de integração de um sequenciador MIDI / Áudio

Interface Gráfico

O *interface* gráfico é na maior parte dos casos, quando usado em conjunto com o tradicional rato e teclado, o meio mais popular de controlo. Fizemos uso deste na acessibilidade à operacionalização do sistema visual (Max/MSP/Jitter). Este ambiente permite a inclusão de elementos de controlo como botões, *faders*, caixas numéricas ou menus, para além de outros bem mais complexos. Fazendo uso destas funcionalidades, implementámos no sistema visual alguns parâmetros de controlo fundamentais. No entanto, tendo em conta a abordagem efectuada a outros métodos de controlo, com uma maior ergonomia e rapidez de acesso, não se sentiu a necessidade de utilizar este de uma forma tão intensiva. É ainda importante referir o facto de, por uma questão de acessibilidade, existirem funções que se encontram expostas simultaneamente em vários métodos de controlo (Ex: botão no *interface* + tecla de função dedicada).

No caso do sintetizador sonoro já referenciado, foi construído o *interface* apresentado anteriormente na Figura 59, pág. 131). Os parâmetros aí presentes controlam directamente a geração sonora e, dada a relação pretendida, indirectamente controlam também o sistema visual. Foi ainda considerada, por motivos de facilidade de operação, a possibilidade de controlo de parâmetros exclusivamente visuais a partir desta aplicação. O painel de controlo denominado de “Video” (Figura 67) encontra-se implementado no Reaktor e poderá conter uma série de opções relativas à configuração do sintetizador visual. Estes parâmetros são enviados por rede através do protocolo OSC, juntamente com os demais parâmetros, para o computador responsável pela síntese de imagem. Se desejável, seria possível eliminar completamente o recurso ao *interface* gráfico do sistema visual no Max/MSP e fazer com que todos os parâmetros de controlo do mesmo (activação e desactivação do metrónomo, da função *fullscreen*, efectuar um *reset* ao sistema, etc.) fossem controlados através deste método.

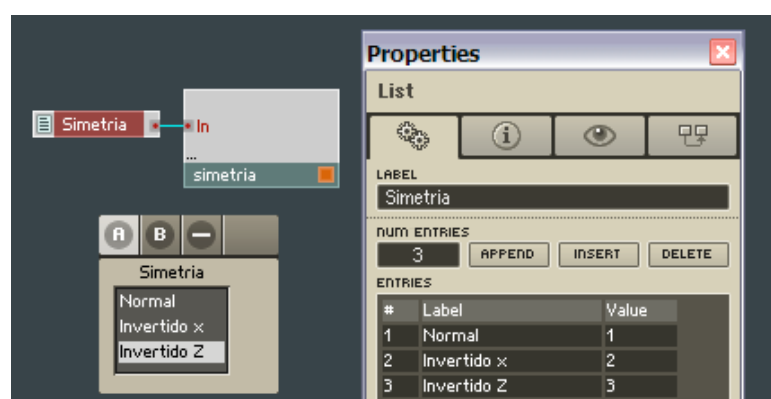


Figura 67 –Estrutura e Painel do *instrument* “Video”, com as propriedades de envio da lista “Simetria” por OSC

Capítulo IV - Conclusão

4.1 Avaliação e análise crítica

No contexto da tarefa que nos propusemos executar será necessariamente difícil construir um método de avaliação capaz de validar ou não o trabalho efectuado baseado na sondagem de opinião de terceiros. Isto porque nunca tivemos a pretensão de sugerir quais as correlações ideais entre os domínios sonoro e visual nem tão pouco de tentar legitimar o valor artístico de determinada obra construída através do sistema proposto. Como afirmado no resumo desta dissertação, concentrámo-nos em efectuar a “*exploração de ferramentas tecnológicas que potenciem a visualização sonora num contexto criativo/artístico de performance musical*”. Deste modo, o recurso a uma avaliação subjectiva por parte de terceiros encontra-se fora do nosso alcance. Em nosso entender, a única forma de aferir o grau de sucesso da implementação sugerida na presente tese será por meio de uma análise à concretização do objectivo formulado inicialmente: a vontade de, através da exploração de uma diversidade de meios tecnológicos, potenciar uma maior correlação dos domínios sonoro e visual num contexto de visualização sonora criativa em tempo-real.

Em nosso entender, a área da visualização sonora encontra-se actualmente dividida em dois grupos. Por um lado, temos os tradicionais **visualizadores de áudio** presentes na maioria dos *media players*, baseados na utilização de algumas características sonoras como o espectro, a forma de onda ou a detecção de ataques, para a geração do resultado visual. Pela análise efectuada no decorrer desta tese, sentimos que, dados os métodos de análise actualmente disponíveis, seria possível e desejável um maior recurso a estes tendo em vista uma maior correlação audiovisual. Por outro lado, consideramos também a existência de uma área de visualização sonora baseada na construção de **soluções personalizadas** para um determinado artista ou espectáculo. Embora, em grande parte dos casos a principal fonte de informação sonora continuar a ser o áudio, o recurso a outras tecnologias como o MIDI já se faz sentir. Em contextos muito particulares de música electrónica ou experimental o recurso ao protocolo OSC é também já uma realidade, motivado sobretudo pela facilitação do transporte da informação (TCP/IP) e pela resolução adicional que este possui, quando comparado com o protocolo MIDI. Constatamos contudo, que estas tecnologias, em particular o OSC não são aproveitadas na sua plenitude.

No decorrer da presente tese, tentámos abordar estas possibilidades tecnológicas com o propósito de extrair o máximo de informação sonora e/ou musical que se possa revelar útil numa correlação com o domínio visual. Em particular, utilizámos a comunicação através do protocolo OSC, para implementar o acesso directo a um conjunto considerável de parâmetros internos de um sintetizador. Pela análise por nós efectuada, julgamos que todas estas abordagens possuem o seu espaço, podendo ser utilizadas em conjunto num contexto de *performance* musical. Cremos que não existe uma única solução que se possa assumir como ideal pelos seguintes motivos:

- Por um lado, **a possibilidade prática de implementação de cada uma destas tecnologias não é de forma alguma universal** uma vez que existem restrições claras no que respeita à sua aplicabilidade. A título de exemplo, não temos acesso directo através do protocolo OSC aos parâmetros internos de geração sonora da maior parte dos instrumentos electrónicos, nem podemos utilizar o protocolo MIDI de forma directa em qualquer tipo de instrumento acústico.
- Por outro lado, **cada uma destas tecnologias possui características complementares**; apesar da necessária etapa de análise, um sinal de áudio digital permite uma caracterização tímbrica dos instrumentos bem mais completa do que o protocolo MIDI que, por sua vez é bastante mais rico ao nível da descrição da *performance* musical. A custo de um maior esforço ao nível da programação, o método de acesso directo aos elementos de controlo da síntese sonora através do protocolo OSC revela-se como o mais completo ao permitir a obtenção simultânea de informações descritivas da *performance* musical e das características tímbricas dos instrumentos.

Na Tabela 11 podemos ver um resumo dos parâmetros sonoros por nós utilizados para uma correlação com o domínio visual. Tendo como premissa inicial a procura de uma maior correlação entre características sonoras e visuais, cremos que a quantidade e a qualidade dos parâmetros disponibilizados são potenciadores de conferir um claro valor acrescentado no âmbito da prática artística da visualização sonora. O facto de não nos limitarmos a um único recurso tecnológico como fonte de informação, aumenta consideravelmente as capacidades do sistema proposto. Para além disso, a ideia de que num contexto de actuação ao vivo é possível obter sinais áudio, MIDI ou OSC relativos a cada um dos instrumentos de forma individual potencia logo à partida uma separação de conteúdos musicais sem a necessidade de recorrer a complexos métodos de análise de áudio. Apesar do nível de correlação obtido através dos parâmetros descritos (Tabela 11), a inclusão adicional de um ou vários “instrumentistas visuais” no sistema proposto vem proporcionar a este uma oportunidade acrescida de colaboração e de expressão artística que julgamos extremamente relevante pois poderá resultar numa mais valia para o resultado final da *performance* audiovisual.

Tendo todos estes aspectos em conta, somos da opinião de que a utilização das técnicas e tecnologias sugeridas no decorrer desta tese potencia uma maior relação audiovisual que, por sua vez, se poderá traduzir num resultado final mais apelativo no âmbito da visualização sonora.

Tabela 11 - Resumo das possibilidades de descrição sonora/musical do sistema proposto

Áudio Digital	
<ul style="list-style-type: none"> Entradas de áudio em tempo-real ou de ficheiros (WAV, AIFF, MP3, AAC), incluindo leitura e correspondente visualização dos <i>tags</i> ID3 (nome, álbum, intérprete, ano), quando aplicável. Possibilidades de representação gráfica da forma de onda do sinal (tempo/amplitude) assim como da sua curva espectral (frequência/amplitude) 	<ul style="list-style-type: none"> Análise do sinal áudio incluindo os seguintes parâmetros: sensação de intensidade (<i>loudness</i>) centro espectral (<i>brightness</i>) quantidade de frequências (<i>noisiness</i>) decomposição espectral segundo a escala de Bark (25 bandas) detecção do ataque de notas posicionamento estereofónico do sinal
MIDI	
<p>Note On / Note Off</p> <p>Program Change</p>	<p>Control Change</p> <p>MIDI Clock</p>
Open Sound Control	
<p>Velocidade de ataque das notas</p> <p>Afinação da voz (polifónico)</p> <p>Amplitude da voz (polifónico)</p> <p>Alteração da afinação (polifónico)</p> <p>Amplitude do LFO relativo ao Filtro</p> <p>Frequência de Corte do Filtro (polifónico)</p> <p>Ressonância do Filtro</p> <p>Amplitude do LFO relativo à Amplitude</p> <p>Amplitude do LFO relativo ao Filtro</p>	<p>ADSR da Amplitude (polifónico)</p> <p>Frequência do LFO1 (polifónico)</p> <p>Frequência do LFO2 (polifónico)</p> <p>Frequência do LFO relativo ao Stereo</p> <p>Amplitude do LFO relativo ao Stereo</p> <p>Nível de Distorção</p> <p>Nível de Chorus</p> <p>Nível de Flanger</p> <p>Frequência do Flanger</p>

4.2 Futuros desenvolvimentos

Analisando o caminho percorrido até agora no âmbito desta tese, quer a um nível de contextualização das tecnologias existentes e documentação daquilo que já foi escrito e construído por outros, quer a um nível mais individual, resultado da formulação e implementação por nós realizada de novos conceitos neste mesmo âmbito, torna-se inevitável uma reflexão relativa a eventuais futuros desenvolvimentos. Somos da opinião de que esta área abordada, dado o seu posicionamento de confluência entre a arte e a técnica, possui a capacidade de gerar muitos pontos de partida para extensões ao trabalho desenvolvido. Pretendemos neste espaço sugerir algumas dessas possibilidades.

- No contexto desta tese foi nosso intuito concentrarmo-nos mais na exploração das possibilidades tecnológicas do que na tentativa de, por meio do sistema proposto, obter um produto artisticamente apelativo. Comprovada a efectividade dos recursos tecnológicos abordados e a viabilidade do sistema proposto, uma possibilidade futura seria a de otimizar o potencial estético do sistema. A este propósito, DeWitt considera a música e a arte visual como as precursoras da *Visual Music*, acrescentando que o estudo das suas estruturas e estéticas tradicionais são fundamentais para a construção desta nova forma de arte [61]. Assim, será necessária uma maior preocupação com alguns aspectos de ordem artística e de percepção visual, como os enunciados por Wilfred [29].
- O protótipo do sistema sugerido, apesar de funcional, é rígido, no sentido em que foi implementado com um objectivo muito particular: o de visualizar simultaneamente um canal de áudio, um instrumento MIDI e um sintetizador sonoro previamente configurado com comunicação por OSC. Não foi propósito no presente contexto realizar uma aplicação flexível e autónoma, passível de ser utilizada em qualquer outra situação com qualquer tipo de *hardware* disponível. Uma possível e desejável extensão do trabalho realizado seria a implementação do necessário grau de modularidade, abstracção e adaptabilidade, bem como de facilidade de operação ao nível do *interface* gráfico, que permitisse e potenciase uma utilização mais generalizada do sistema sugerido.
- A implementação de métodos de análise de áudio e MIDI mais evoluídos para a discretização do discurso musical a níveis semânticos mais altos, como a análise musical, permitiria que pudessem ser mapeados parâmetros como a forma musical e motivos melódicos e/ou rítmicos a parâmetros de geração de imagem também de alto nível como formas ou texturas complexas.
- No decorrer do trabalho proposto tivemos como propósito representar visualmente os eventos sonoros, seja esta representação encarada como a origem sonora ou mesmo como a sua consequência. Podemos considerar que foi procurada uma relação harmónica entre som e imagem. No entanto, seria possível e até desejável a existência conjunta de uma abordagem alternativa, uma

espécie de contraponto entre o domínio sonoro e o visual. Michel Chion [6] refere que o contraponto audiovisual fomenta o desvio de uma certa redundância entre som e imagem, para uma forma onde estes possam ser encarados como dois elementos paralelos ligados de uma forma pouco rígida, não estando sempre um dependente do outro. Grande parte das obras musicais combina de algum modo estas duas dimensões, harmonia e contraponto. Extrapolando esta técnica composicional para o domínio audiovisual, poderemos considerar os elementos visuais não apenas como meras representações dos eventos sonoros, mas como instrumentistas autónomos que “ouvindo” a música, contribuem activamente para o produto audiovisual.

- Uma abordagem complementar a este trabalho poderia advir da ideia de que o som possa também ser alterado em função de determinados parâmetros da imagem gerada, formando assim um sistema de *feedback*. Podemos argumentar que esta ideia se encontra aqui presente de uma forma muito indirecta: dado que o sistema pretendido funciona em tempo-real, o próprio músico pode ser influenciado na sua actuação pela visualização resultante. Esta influência potencialmente poder-se-á concretizar em opções musicais que na ausência deste estímulo visual não seriam tomadas.

Pela análise ao trabalho por nós realizado e pela projecção de possíveis futuros desenvolvimentos como os aqui sugeridos, somos da opinião de que existe claramente matéria para uma eventual continuação de investigação e exploração neste mesmo âmbito. Conforme referido anteriormente, independentemente da validação de possibilidades de correspondência entre características sonoras e visuais, tão debatida ao longo do tempo por teóricos e artistas, somos da opinião que o simples acto criativo de exploração e combinação destas constitui à partida uma oportunidade de expressão artística extremamente relevante.

Bibliografia

- [1] H. M. a. S. Garden. (2005, *Visual Music*. Available: <http://hirshhorn.si.edu/visualmusic/>
- [2] F. Collopy, "Color, Form and Motion: Dimensions of a Musical Art of Light," *Leonardo*, vol. 33, 2000.
- [3] A. B. Klein, *Color Music: the art of light*. London: C. Lockwood and son, 1927.
- [4] W. Moritz, *Optical Poetry: The Life and Work of Oskar Fischinger*: Indiana University Press, 2004.
- [5] K. Brougher, *et al.*, *Visual music : synaesthesia in art and music since 1900*. Los Angeles, CA: Museum of Contemporary Art, Los Angeles, 2005.
- [6] M. Chion, *et al.*, *Audio-vision : sound on screen*. New York: Columbia University Press, 1994.
- [7] P. Friedlander. *Paul Friedlander*. Available: <http://www.paulfriedlander.com/text/visualmusic.html>
- [8] S. D. Lipscomb and E. M. Kim, "Perceived match between visual and auditory correlates: an experimental multimedia investigation," presented at the International Conference on Music Perception & Cognition (ICMPC8), 2004.
- [9] D. Gerhard, *et al.*, "Exploration of the correspondence between visual and acoustic parameter spaces," presented at the New Interfaces For Musical Expression, Hamamatsu, Shizuoka, Japan, 2004.
- [10] J. Cloutier, *A era de EMEREC ou a comunicação audio-scripto-visual na hora dos self-media*. Lisboa: Instituto de Tecnologia Educativa, 1975.
- [11] L. L. Henrique, *Acústica Musical*: Fundação Calouste Gulbenkian, 2002.
- [12] T. D. Rossing, *The Science of Sound*: Addison-Wesley Publishing Company, 2001.
- [13] D. H. Hubel, *Eye, Brain, and Vision*: W. H. Freeman, 1995.
- [14] K. Peacock, "Instruments to Perform Color-Music: Two Centuries of Technological Experimentation," *Leonardo*, vol. 21, 1988.
- [15] J. W. v. Goethe, *Theory of Colours*: The MIT Press, 1970.
- [16] F. Collopy. *Rhythmic Light*. Available: <http://www.rhythmiclight.com/>
- [17] A. W. Rimington, *Colour-Music: The Art of Mobile Colour*: Adamant Media Corporation, 2002.
- [18] W. Kandinsky. (1912). *Concerning The Spiritual in Art*. Available: <http://www.gutenberg.org/etext/5321>
- [19] T. D. Jones, *The Art of Light and Color*: Van Nostrand Reinhold, 1972.
- [20] F. Popper, *Origins and Development of Kinetic Art* Littlehampton Book Services Ltd, 1968.
- [21] K. Gerstner, *The Forms of Color*: The MIT Press, 1990.
- [22] R. Arnheim, *Art and Visual Perception: A psychology of the Creative Eye*: University of California Press, 2004.
- [23] G. Kepes, *Language of Vision*: Dover Publications, 1995.
- [24] W. Kandinsky, *Point and Line to Plane*: Dover Publications, 1979.
- [25] L. E. Marks, *The Unity of the Senses: Interrelations Among the Modalities*: Academic Press, 1978.
- [26] M. Graves, *The Art of Color and Design*: McGraw Hill Book Company, Inc., 1941.
- [27] O. N. Rood, *Modern chromatics; students' text-book of color, With applications to art and industry* Kessinger Publishing, 2008.
- [28] R. Scruton, *The aesthetics of music*. Oxford ; New York: Oxford University Press, 1999.
- [29] T. Wilfred, "Light and the Artist," *The Journal of Aesthetics & Art Criticism*, vol. V, 1947.

- [30] G. Levin and Massachusetts Institute of Technology. Dept. of Architecture. Program In Media Arts and Sciences., "Painterly interfaces for audiovisual performance," Thesis S.M. --Massachusetts Institute of Technology Program in Media Arts & Sciences 2000., 2000.
- [31] O. Wilde, *Essays and Lectures*. New York: G. P. Putnam's Sons, 1894.
- [32] W. Pater. (1877). *The Renaissance: Studies in Art and Poetry*.
- [33] H. Cristol and K. Hileman, "Exhibition traces music as an inspiration for visual art in the 20th century," *Inside Smithsonian Research*, 2004.
- [34] J. Whitney, *Digital Harmony: On the Complementarity of Music and Visual Art*: McGraw-Hill Inc., 1980.
- [35] J. Watkinson, *An Introduction to Digital Audio*: Focal Press, 1994.
- [36] A. P. Kefauver, *Fundamentals of Digital Audio*: A-R Editions, 1999.
- [37] *Understanding Metadata*: National Information Standards Organization Press, 2004.
- [38] H.-G. Kim, *et al.*, *MPEG-7 audio and beyond - audio content indexing and retrieval*. Chichester, West Sussex, England ; Hoboken, N.J.: John Wiley and Sons, 2005.
- [39] M. E. P. Davies, *et al.*, "Beat Tracking Towards Automatic Musical Accompaniment," Barcelona, Spain, 2005.
- [40] P. M. Brossier, *et al.*, "Fast labelling of notes in music signals," 2004.
- [41] V. Sandvold and P. Herrera, "Towards a Semantic Descriptor of Subjective Intensity in Music," in *International Computer Music Conference*, Barcelona, Spain, 2005.
- [42] A. Berenzweig, *et al.*, "Using Voice Segments to Improve Artist Classification of Music," presented at the International Conference on Virtual, Synthetic and Entertainment Audio, Espoo, Finland, 2002.
- [43] Y. E. Kim and B. Whitman, "Singer Identification in Popular Music Recordings Using Voice Coding Features," in *3rd International Conference on Music Information Retrieval*, Centre Pompidou - Paris, France, 2002.
- [44] S. Essid, *et al.*, "Efficient musical instrument recognition on solo performance music using basic features," presented at the AES 25th International Conference, London, United Kingdom, 2004.
- [45] E. Gómez and P. Herrera, "Estimating the tonality of polyphonic audio files: cognitive versus machine learning modelling strategies," in *5th International Conference on Music Information Retrieval*, Barcelona, Spain, 2004.
- [46] B. Martin, *et al.*, "Musical structure retrieval by aligning self-similarity matrices," in *ISMIR 2009*, Kobe, Japan, 2009.
- [47] D. Liu, *et al.*, "Automatic Mood Detection from Acoustic Music Data," presented at the ISMIR 2003, Baltimore, Maryland, USA, 2003.
- [48] *ITMA MIDI Detailed Specification*: International MIDI Association, 1990.
- [49] P. D. Lehrman and T. Tully, *MIDI for the Professional*. New York: Amsco Publications, 1993.
- [50] M. Wright. (2002, *OpenSound Control Specification*. Available: <http://archive.cnmat.berkeley.edu/OpenSoundControl/OSC-spec.html>
- [51] F. Fluckiger, *Understanding networked multimedia : applications and technology*. London ; New York: Prentice Hall, 1995.
- [52] M. Wright, *et al.*, "Open Sound Control: State of the Art 2003," presented at the International Conference on New Interfaces for Musical Expression, Montreal, 2003.
- [53] M. Russ, *Sound synthesis and sampling*. Oxford ; Boston: Focal Press, 1996.
- [54] S. Crab. (2005, *120 Years of Electronic Music*. Available: <http://120years.net/>

- [55] J. Chowning, "The Synthesis of Complex Audio Spectra by Means of Frequency Modulation," *Journal of the Audio Engineering Society*, 1973.
- [56] C. Roads, *The Computer Music Tutorial*: The MIT Press, 1996.
- [57] J. D. Foley, *et al.*, *Introduction to Computer Graphics*: Addison-Wesley Publishing Company, 1994.
- [58] C. Wootton, *A practical guide to video and audio compression : from sprockets and rasters to macroblocks*. Amsterdam ; Boston: Elsevier Focal Press, 2005.
- [59] M. Woo, *et al.*, *OpenGL programming guide : the official guide to learning OpenGL, version 1.1*, 2nd ed. Reading, Mass.: Addison Wesley, 1997.
- [60] R. J. Rost, *et al.*, *The OpenGL Shading Language*, 2nd ed. Upper Saddle River, NJ: Addison-Wesley, 2006.
- [61] T. DeWitt, "Visual Music: Searching for an Aesthetic," *Leonardo*, vol. 20, pp. 115-122, 1987.
- [62] J. C. Whitaker and K. B. Benson, *Standard Handbook of Video and Television engineering*, 4th ed. New York: McGraw-Hill, 2003.
- [63] A. S. Tanenbaum, *Computer networks*. Englewood Cliffs, N.J.: Prentice-Hall, 2003.
- [64] T. Jehan and Massachusetts Institute of Technology. Dept. of Architecture. Program in Media Arts and Sciences., "Creating music by listening," Thesis Ph. D. --Massachusetts Institute of Technology School of Architecture and Planning Program in Media Arts and Sciences 2005., 2005.
- [65] D. Wessel, "Max/MSP Programming Practice with OSC," presented at the Open Sound Control Conference, Berkeley, CA, 2004.
- [66] D. Gibson, *The Art of Mixing: A Visual Guide to Recording, Engineering and Production*: MixBooks, 1997.